The Journal of
Applied Psychology

Edited by
JAMES P. PORTER
Ohio University
WILLIAM F. BOOK
Indiana University
With a Board of Co-operating Editors

VOLUME VII
1923

PUBLISHED BY THE EDITORS

Reprinted with the permission of the original publishers.
JOHNSON REPRINT CORPORATION
KRAUS REPRINT CORPORATION

The Journal of Applied Psychology

CO-OPERATING EDITORS

W. V. Bingham, Carnegie Institute of Technology; June E. Downey, University of Wyoming; Robert H. Gault, Northwestern University; William Healy, Baker Foundation, Juvenile Court, Boston; H. L. Hollingworth, Columbia University; Joseph Jastrow, University of Wisconsin; Herbert S. Langfeld, Harvard University; James H. Leuba, Bryn Mawr College; Henry C. McComas, Princeton University; Walter D. Scott, Northwestern University; C. E. Seashore, *University of Iowa*; Daniel Starch, University of Wisconsin; Edward K. Strong, Jr., Carnegie Institute of Technology; L. M. Terman, Leland Stanford Jr. University; E. L. Thorndike, Columbia University; J. E. Wallace Wallin, Teachers College, Miami University; G. M. Whipple, University of Michigan; Helen Thompson Woolley, Vocation Bureau, Cincinnati.

The scope of the Journal of Applied Psychology will include the following: (a) The application of psychology to vocational activities, such as law, art, public speaking, industrial and commercial work, and problems of business appeal. (b) Studies of individual mentalities, such as types of character, special talents, genius, and individual differences, including the problems of mental diagnosis and vocational prognosis. (c) The influence of general environment conditions, such as climate, weather, humidity, temperature; also such conditions as nutrition, fatigue, etc. (d) The psychology of everyday activities, such as reading, writing, speaking, singing, playing games or musical instruments, sports, etc.

TABLE OF CONTENTS

Meta Anderson

A study of the data on the results gathered from repeated mental examinations of 200 defective children attending special schools over a period of eight years 54-64

Ada Hart Arlitt

Intelligence tests versus entrance examinations as a means of predicting success in college 330-338

Coryne Basch and Harold E. Burtt

Legibility of Bodoni, Baskerville Roman, and Cheltenham Type Faces 237-245

Marian A. Bills

Relation of mental alertness test score to positions and permanency in company 154-156

Harold E. Burtt and J. Camden Clark

Facial expression in advertisements 114-126

Harold E. Burtt and F. W. Ives

Vocational tests for agricultural engineers 178-187

Harold E. Burtt and Coryne Basch

Legibility of Bodoni, Baskerville Roman, and Cheltenham Faces 237-245

J. Camden Clark and Harold E. Burtt

Facial expressions in advertisements 114-126

Laura C. Craytor and C. B. Davenport

Comparative social traits of various races 127-134

J. F. Dashiell and John F. McFadden

Racial differences as measured by the Downey Will-Temperament Test 30-53

C. B. Davenport and Laura C. Craytor

Comparative social traits of various races . . . 127-134

M. C. Del Manzo and G. M. Ruch

The Downey Will-Temperament Group Test: a further analysis of its reliability and validity . 65-76

Gustave A. Feingold

The relation between the intelligence and vocational choices of high school pupils 143-153

William S. Foster

Experiments on rod-divining 303-311

R. B. Franken and A. T. Poffenberger

A study of the appropriateness of type faces . . . 312-329

Arthur I. Gates

The unreliability of M. A. and I. Q. based on group tests of general mental ability 93-100

B. F. Haught

A scheme for combining incomplete rankings . . . 168-172

C. L. Huffaker

A contribution to the technique of partial correlation 135-142

Clark L. Hull

Prediction formulae for teams of aptitude tests . 277-284

F. W. Ives and Harold E. Burtt
Vocational tests for agricultural engineers 178-187
Grace H. Kent
A combination mental test for clinical use 246-257
H. D. Kitson, C. W. Barnes and C. Iuppenlatz
Minor studies in the psychology of advertising . 10-15
F. B. Knight and H. H. Remmers
Fluctuations in mental production when motivation is the main variable 209-223
Arthur W. Kornhauser
A re-interpretation of the statistical method of Army Trade Tests 339-348
Donald A. Laird
The basis of tooth paste sales in representative communities 173-177
John H. McFadden and J. F. Dashiell
Racial differences as measured by the Downey Will-Temperament Test 30-53
Blanche M. Minogue
A case of secondary mental deficiency with musical talent 349-352
Elsie Murray
Freshman tests in the small college 258-276
Everett F. Patton
An experiment in testing engine lathe aptitude . 16-29
A. T. Poffenberger
The conditions of belief in advertising 1-9
The return coupon as a measure of advertising efficiency 202-208
Measuring the comprehension of advertisements 364-369
A. T. Poffenberger and R. B. Franken
A study of the appropriateness of type faces . . 312-329
H. H. Remmers and F. B. Knight
Fluctuations in mental production when motivation is the main variable 209-223
W. T. Root
The freshman: Thorndike College Entrance Tests, first semester grades, Binet Tests 77-92
G. M. Ruch and M. C. Del Manzo
The Downey Will-Temperament Group Test: a further analysis of its reliability and validity . . 65-76
Irene Case Sherman
The Franz Dot Tapping Test as a measure of attention 353-359
A. J. Snow
An advanced course in experimental psychology applied to advertising 360-363
Labor turnover and mental alertness test scores . 285-290
M. LaVinia Warner
Influence of mental level in the formation of boys' gangs 224-236
J. U. Yarbrough
A study of the correlation of college students' estimates of intelligence with the Otis and other scales 157-167

THE
JOURNAL OF APPLIED PSYCHOLOGY

VOL. VII MARCH, 1923 No. 1

THE CONDITIONS OF BELIEF IN ADVERTISING

A. T. Poffenberger
Columbia University

A reader of the literature dealing with the Psychology of Advertising finds numerous articles written upon the effectiveness of various mechanical devices, the attention and memory value of size of space and the position on the page, the influence of color, style of type and its arrangement, the effectiveness of repeating the advertisement, etc. The problem of arousing the *confidence of the consumer in the article advertised*, the conditions on which it depends, how belief in advertisements may be created and how it may be measured have been very lightly touched in experimental studies. The importance of these matters is emphasized in every textbook on advertising. The fact that the American people are each year induced to squander many millions of dollars in worthless securities through the medium of advertising in some form, and that warnings seem quite ineffective in protecting them, makes one curious about the basis of belief in advertising. It is not enough to say that the American people like to be fooled and that there is no scheme too wild to arouse the confidence of a large proportion of them. The advertiser should know that action is dependent upon belief and that belief in advertising depends upon conditions, some of which at least are under his control.

Belief is indeed a complex mental state and it may depend at any time upon a great variety of factors, the most common of which are listed in textbooks of psychology and advertising. It is not the purpose of this report to repeat such lists but to cite some experimental evidence from the field of advertising to show the importance of a few of the conditions of belief for success in advertising.

1. Belief is rarely the result of reasoning. One does not go through the processes of logic to establish his beliefs. If logic is used at all it is to justify a belief already established. A strik-

ing illustration of the separation of reasoning and belief is found in the case of the insane patient who firmly believed himself to be the son of a king, and yet whose reason was intact enough to enable him to solve complicated mathematical problems.

The advertising of the New Gillette razor offered a good opportunity for studying the relation between belief and reasoning.[1] On May 16, 1921, the Gillette Razor Company announced "a new triumph of American inventive genius of startling interest to every man with a beard to shave." The advertisements state that the "fulcrum shoulder, overhanging cap and channeled guard" were the three innovations which made possible "for the first time in any razor, micrometric control of blade position." A diagram showed "how the blade is biflexed between overhanging cap and fulcrum shoulder. It is flexed once into the inside curve of the cap. This is the minor flexure—the curve for easy gliding action and play of the wrist in shaving. It is flexed a second time—more sharply and in a shorter radius—by the grip of the overhanging cap the whole length of the fulcrum shoulder. This is the major flexure." This arrangement provided an exactness of adjustment to 1-1000 of an inch. Advertisements containing the above information and well illustrated were given to fifty-seven men, college students and university graduates, together with a series of seven questions intended to test both their belief in the new razor and their understanding of it. The answers to these questions showed that all the students agreed that the new razor was better than the old one, and that they would rather pay \$5.00 for the new one than \$1.00 or \$2.00 for the old one. In supporting their belief they were allowed to consult the advertisement as much as they wished. They quoted the "fulcrum shoulder, overhanging cap and channeled guard," which made possible "micrometric control of blade position," but not one of them could explain how the micrometric control was obtained or what advantage there would be in having such micrometric control. They believed that the "channel guard" was an improvement although they could not tell why it was an improvement. As to the importance of major and minor flexures they were entirely ignorant. Five minutes examination of an enlarged diagram of the new razor improved their understanding of the razor little or not at all. Here is a belief effective for the purpose of the advertiser in spite of the inability of the reader to support his belief with reason. This experiment is quoted not to show that the advertisement is poor, but merely to show that in advertising reasoning is not needed to create belief. Whether the space in the advertisement devoted to argument might have been more profitably filled, is however an interesting question.

---

[1] The study about to be described is part of a larger investigation conducted by David Goldstein and the writer, which will be reported at a later date.

2. It is not necessarily that which is true that is believed. The fact that a statement in an advertisement is true will not guarantee belief on the part of the readers of it. The truth may be too startling and surprising to be believed, and in some cases it might be more effective to tell half the truth than the whole truth. Three advertisements were chosen for investigation in this connection, each of them presenting rather unusual information. In each case the truth of the statements made was vouched for by reliable individuals. Each of the advertisements was presented to 100 people, together with a series of statements about them. The individuals were to simply check those statements that most nearly represented their opinion about the advertisements. The statements were carefully prepared so as to avoid suggestion. The studies of the three advertisements will be reported separately:

Taylor Trunk Advertisement. This advertisement showed the photograph of a huge elephant standing on a trunk. The picture was accompanied by a signed statement as to the genuineness of the photograph, and by statements to the effect that the trunk was taken from the regular stock. Statistical treatment of the replies of the persons tested showed that 38 percent of them doubted the truth of the statements made in the advertisement; 24 per cent questioned the genuineness of the photograph; and 21 percent believed that it would be impossible to construct a trunk strong enough to withstand such a weight. That is, in from one-fifth to two-fifths of the persons tested the advertisement created a state of mind adverse to the purchase of the article.

Stanley Vacuum Bottle Advertisement. This advertisement represented a vacuum bottle falling from a high window, and described an incident in which a Stanley bottle had accidentally fallen from an eighth story window and suffered only a dent, which did not at all interfere with its usefulness. The replies to the questions submitted with this advertisement showed that 31 percent doubted the truth of the statements, while 22 percent refused to believe that a vacuum bottle sufficiently strong to survive the fall could possibly be constructed.

Edison Phonograph Advertisement. This advertisement pictured one of the well known "tone tests" demonstrations in Carnegie Hall, Pittsburg. It showed the stage occupied by a well known singer and an Edison phonograph, with the audience in the background. On the advertisement there were reproduced a number of clippings from the daily papers describing the remarkable demonstration, each clipping making the statement in one form or another that no one in the audience of 2,000 people could distinguish between the voice of the singer and its reproduction. Replies from 100 people to eight questions about this advertisement may be summed up as follows: 77 percent doubted the truth of the statements contained in the advertisements. 73 percent believed that they could tell the difference between a real voice and its reproduction by any phonograph. 82 percent believed that mechanical sounds would betray the phonograph in these "tone tests." 49 percent believed that some form of trickery or deception was practiced. 50 percent believed that the artist intentionally imitated the phonograph. 44 percent believed that a specially constructed phonograph was used for the demonstration rather than a stock instrument. 68 percent believed that no phonograph could successfully undergo such a test. 39 percent stated that an actual demonstration of the "tone test" such as that described in the advertisement would not convince them of their inability to distinguish between the human voice and its reproduction.

This advertisement seemed especially effective in failing to create belief in the minds of its readers, when over three-fourths of them doubted the truth of the statements. The writer recommended in this case a less extreme and less startling form of appeal which should not arouse the antagonism manifested in this experiment.

Thus far the negative side of the question has been presented. If belief in an advertisement does not depend upon the truth of the statements made and does not depend upon the reasoning of the reader, on what does it depend? To state the matter simply, we may say that ideas which are present in the mind and are not interfered with by any opposing ideas will be believed. This is merely a bare statement of the law of suggestion and to comply with it in advertising, conflicting ideas should be prevented from entering the mind. There are many conditions on which such undisturbed acceptance of ideas depends. Only three will be mentioned here.

1. The ideas aroused by an advertisement must not conflict too sharply with the reader's experience. Introspections volunteered by many of the 100 subjects who served in the experiments just described indicated that their past experiences with trunks, vacuum bottles and phonographs furnished conflicting ideas which the advertisements were not sufficiently powerful to overcome. This was especially true in the case of the phonograph advertisement where doubt was expressed in a large percentage of the cases.

An experimental study of an advertising campaign that failed showed clearly the need for complying with the conditions of belief. "Radior Products," a series of toilet preparations were introduced into this country by an English firm. The appeal contained in the advertising may be illustrated by the following quotation taken from one of the advertisements: "Radior is the magic new word in the book of beauty. It means the triumphant union of the finest complexion preparations with actual radium. Its content of radium works the miracle of nature. It purifies the skin, gives it the health to regain its youthfulness and loveliness." The container of these preparations was represented as emanating rays which were very suggestive of bolts of lightning. These toilet preparations could not be sold—the campaign had failed. Why? One possible reason for the failure was that the suggestion to apply radium on the skin for toilet purposes conflicted too sharply with people's opinions about radium. In order to discover what people believed about radium, a questionaire was carefully prepared in the form of a True-False test and submitted to 400 people, comprising a well-to-do, well-educated group, a group of average intelligence and financial standing and a special group of persons working in "Beauty Parlors." The questionaire is reproduced below. It was constructed so as to minimize the influence of suggestion and to correct for it when it did occur, as for instance in combining replies from question 2 and 7. Questions 3, 5 and 10 were introduced as fillers and had no special significance.

Read the first statement on this sheet. If you *feel* that it is true, underline the word "True." If you feel that it is false, underline the word "False." Then judge each of the other statements as true or false. If you do not know, guess at it. This sample statement is false and is marked as it should be:

There are eight days in the week..........................True False

1—Radium is a deadly poison....................................True False

2—Radium is so costly that it cannot be
used for commercial purposes.................................True False

3—Treatments with radium will make
the hair grow ....True False

4—Radium is used in the treatment of
such diseases as cancer..True False

5—Radium is a substance which gives off light......True False

6—Radium when applied in diluted form gives
a healthful stimulation to the skin.........................True False

7—Radium is now used in the manufacture
of many products in common use..........................True False

8—Radium causes burns when it comes into
contact with the skin...True False

9—Toilet preparations containing radium are to be recommended for their beneficial effects........True False
10—Certain forms of radium can be obtained at little cost ....True False
11—Radium preparations should not be used except on the advice of a physician......................True False
12—Radium products are gaining popularity as mild stimulants to health and growth..................True False

Read the following question and answer it by underlining "Yes" or "No":

If you should be told that there was a toilet preparation of high quality and moderate price and which contained a minute quantity of radium, would your knowledge of the effects of radium be such as to make you want to buy it?..................................Yes No

Nothing need be said here concerning the difference in the reactions of the different groups to the questionaire, except that the most intelligent group had the strongest reaction against radium-containing products, and that the "Beauty Parlor" workers had the least antagonism. Inquiry among these people indicated that some of them had confused radium with the violet ray which was at the time a very popular form of treatment with them. The results of the study of the questionaire may be summarized as follows:

Forty-eight per cent of the persons tested believed that radium is a deadly poison; 80 per cent believed that it causes burns when it comes in contact with the skin; 89 per cent believed that radium preparations should be used only upon the advice of a physician; 71 per cent said that they would not buy any kind of toilet preparation that they knew contained even a minute quantity of radium; 90 per cent thought of radium as a substance used for the treatment of cancer. The whole experiment may be summed up in the statement that about 70 per cent of all the replies indicated opinions unfavorable to radium-containing products as toilet preparations. The basis of this unfavorable reaction could easily be traced to the people's accumulated experiences of radium. Clippings of all articles dealing with radium and appearing in newspapers and magazines were collected for a short period. Practically all of them emphasized the harmful effects of radium and the dangers incurred in handling it, instead of any beneficial properties it might possess.

This is a clear case of the inability of an advertising campaign to overcome the resistance established by experience. To create a favorable attitude toward radium-containing products as toilet preparations by an educational campaign conducted in newspapers and magazines might conceivably have been possible. But it certainly would have been impracticable.

The mass of experiences upon which beliefs rest is gradually accumulating and being modified by additions. As some one has said, "Confidence is at the same time stable and fragile." It is stable because of the bulk of experiences on which it rests; it is fragile because it is subject to change from new experiences. A supposed bargain which is found to be not a bargain, a derogatory newspaper report about a store, a misprint in the advertised price of an article; any one of these may be the experience that will introduce a conflicting idea into the mind of a consumer and upset the conditions favorable for his belief. The advertiser, appealing as he does to populations, cannot take account of the shiftings of individual experience and belief, but he must take account of the widespread and slowly accumulating experiences which may amount to powerful prejudices against his appeal.

2. Ideas that are to create belief must come from an authoritative source. This is a well-known law of suggestion. The hypnotist can do nothing without his air of authority and the subject's recognition of it. We are accustomed to believe the statements made by a person in whom we have confidence, and to believe what is printed in a medium which we consider authoritative. Even if there is conflict with one's own experience, he will sometimes accept the contrary experience of another person as a basis for belief if he have sufficient confidence in the other person. But even then the new experience may not be too conflicting. Advertisers have for years striven to develop an atmosphere of confidence and authority by all the devices at their command. The present experiment was intended to measure in a tentative fashion the degree of confidence which an advertisement can create in comparison with other forms of publication. I have compared the degree of belief or doubt aroused by the three advertisements previously described (namely Taylor Trunks, Stanley Vacuum Bottles and Edison Phonographs) with the belief aroused by essentially the same statements coming from a reputable journal. In order to make such a comparison, the facts stated in each advertisement were prepared in the form of a news item abstracted from an engineering magazine. One such abstract will illustrate the character of all. These abstracts were presented to a group of 100 persons of the same general character as those tested in the earlier experiments, but who knew nothing of those experiments or nothing of the purpose of the present experiment. Along with the abstract of each advertisement was a series of questions as nearly as possible like those used in the test with the advertisements. As far as we were able to ascertain, no one doubted the

authenticity of the news abstracts. Instead of reporting the results of the three questionaires in detail, it will be sufficient to compare for each of the three cases the number of replies indicating doubt in the advertisement and doubt in the news item. It must be remembered that the news item mentioned no trade-named article nor indicated any connection with specially advertised goods.

The following is an extract from——————Magazine of June, 1921. Read it carefully and then answer the questions given below, by scratching out with a pencil the part of each that does *not* give your opinion.

A Remarkable Shock Resisting Construction

"A new type of wardrobe trunk was recently put through a remarkable test of strength. The five-ton elephant belonging to the New York Hippodrome was made to stand on the empty trunk. An examination of the trunk following the test showed that it had withstood the great shock without damage.

Interesting photographs were made of the test, showing the elephant mounting the trunk, seated upon it, and then stepping off of it."

1—I believe—doubt—do not believe—the statements made in the above extract.

2—I feel that the photographs that portray the test are—are not—genuine.

3—I do—do not—believe that a trunk could be made strong enough to successfully undergo the test described.

Percentage of replies expressing doubt	*In the Ads*	*In the News Items*
Taylor Trunks	28 per cent	43 per cent
Stanley Vacuum Bottles	27 per cent	43 per cent
Edison Phonographs	55 per cent	40 per cent

In two of the cases it will be noted that there was greater confidence in the advertisement than in the news clipping, while in the third there was greater confidence in the news clipping. An examination of the three advertisements did not afford an entirely satisfactory explanation for the shift of belief in the case of the third advertisement. It seems safe to conclude from these records that although belief in certain advertisements may be low, they may carry at least as much authority in presenting a set of facts as can be conveyed by a news article. The doubt in the cases we have studied is the effect rather of conflict of ideas with experience, than the effect of the use of an unauthoritative medium of expression. Introspections volunteered by the subjects suggest that illustrations and especially photographs used in the advertisements tend to strengthen belief. Unless one suspects trickery, as some of our subjects did, looking at a picture ought to carry with it greater weight in establishing belief than merely reading printed matter. Even if Mark Twain was right in advising that one believe only half that he sees and nothing that he hears, the advantage in favor of the picture is obvious.

3. There is a third important condition of belief, namely, that we tend to believe what arouses our desires, our fears and our emotions generally. I have no experimental evidence to offer in this connection, and know of none in the field of advertising. But evidence for the importance of this factor may be drawn from psychology. Wm. James has said, "A man who has no belief in ghosts by daylight will temporarily believe in them when alone at midnight, he feels his blood curdle at a mysterious sound or vision, his heart thumping and his legs impelled to flee." In strong emotion we might find the condition responsible for the belief in the bargain counter. The politician finds no difficulty in honestly believing what best fits in with his aspirations, while his opponent may as honestly believe the opposite and for a like reason. If one really wants a certain suit of clothes or an automobile which costs more than he should pay, he may honestly believe that he is making an economical purchase. Once a belief has been established in this way, logic and reasoning may be used to support it.

These three conditions of belief which I have described represent three important variables in the formula for advertising success. For most advertising situations they are unknown variables which may interact in a very complex manner. But they may be discovered by careful examination. That they are not always sought or discovered, is clear from the cases I have cited which were taken from advertising already used. Three questions might well be asked about every advertisement: (1) What adverse beliefs and experiences does it have to meet in the minds of consumers? (2) Will the authority which it wears by its mode of presentation or by the medium in which it appears enable it to create belief? (3) Does the appeal used arouse desires which will in turn create belief in the advertised article? These are human behavior questions that psychological methods will enable the advertiser to answer before the advertising is used as well as after the money has been spent upon broadcasting it.

MINOR STUDIES IN THE PSYCHOLOGY OF ADVERTISING

From the Psychological Laboratory of Indiana University

By H. D. Kitson, C. W. Barnes, and C. Iuppenlatz

IX. Right and Left Hand Pages in Magazines

A considerable amount of discussion has centered around the question as to which is the more favorable position for advertisements in magazines, the right or left-hand page. The question has usually been approached through the laboratory method, subjects viewing the momentarily exposed pages of a "dummy" or a real magagine and then writing down all they have seen in the order in which they have seen it. With this technic Adams[1] reports a value of the right hand page which is to that of the left as 100 is to 65. That is, he estimates the right hand to be fifty per cent superior. Hotchkiss and Franken[2] estimate from their investigation that the right-hand page of a magazine is worth five per cent more than the left-hand page.

In order to secure additional light upon the question the writers approached the problem with the historical technic, seeking to discover the frequency with which advertisers actually used the right-hand and left-hand pages.

The materials used were 1082 consecutive pages of advertising matter in the American Magazine and 1000 in the Saturday Evening Post, for the year 1922. Beginning with the first advertisement following solid reading matter tabulation was made of the successive pages containing advertising matter, exclusive of cover pages. Results are shown in Table I.

TABLE I.

SHOWING THE FREQUENCY WITH WHICH RIGHT AND LEFT-HAND PAGES CARRY ADVERTISEMENTS IN THE AMERICAN MAGAZINE AND THE SATURDAY EVENING POST FOR 1922

	Total Right		Total Left		Total Total	Full Right		Full Left		Full Total
	No.	%	No.	%		No.	%	No.	%	No.
Amer.	576	56	506	44	1082	301	82	64	18	365
S. E. P..........	521	52	481	48	1000	350	67	171	33	521

---

[1] H. F. Adams, Advertising and its Mental Laws, New York, Macmillan, 1916. p. 100.

[2] G. B. Hotchkiss and R. B. Franken, Attention Value of Advertisements, New York, New York University, 1920, p. 24.

These figures show that in both magazines the right-hand page is used for advertisements more frequently than the left; in the American 56 per cent of the times and in the Saturday Evening Post 52 per cent of the times; with a frequency on the average 17 per cent greater than that for the left-hand page.

Still stronger evidence of the favor in which the right-hand page is held comes from facts shown in a separate tabulation which was made concerning the location of full page advertisements. There were 365 of these in the American Magazine and 521 in the Saturday Evening Post. Eighty-two per cent of the former and 67 per cent of the latter appeared on the right-hand page. In the first named periodical the right-hand page was used for full page advertisements five times as frequently as the left, and in the Saturday Evening Post twice as frequently. It was observed that certain advertisers who consistently used the same relative position in issue after issue almost invariably used the right-hand page.

On the assumption of the historical method, that a practice much in vogue has probably proved itself profitable, it would appear from these results that the right-hand page in magazines is more profitable than the left. True, the practice here revealed need not necessarily be considered as having been based upon conscious reckoning of comparative profits and losses from right and left-hand insertions. Even at best it probably represents only the fruits of a "sub-conscious" "hunch." Nevertheless it is coming to be recognized that such a process of summing up experience has some validity as a guide in business practice where scientifically valid measures are not yet obtainable.

Appearance on the right or left-hand page of a magazine should, however, not be considered as wholly a result of the voluntary choice of the advertiser. Requirements of journalistic make-up would inevitably put some advertisements on the right and some on the left-hand page. Again, the requirements of color process or of some other technical matter might force certain advertisements on one page rather than another. Moreover, some advertisers do not specify either page, not regarding the matter as important enough to bother with. For these reasons the more frequent use of the right-hand page can not be regarded as purely a reflection of profitable experience. Still these factors alone might be expected to work toward an equal distribution of advertisements upon the right and left-hand pages. The fact that they are unequally distributed indicates that some selective factor is at work in favor of the right-hand page. And the fact that this page is used with overwhelming frequency by the purchasers of the expensive full page advertisements indicates that the right-hand page has demonstrated its superiority to at least some influential and successful advertisers. On slighter evidence than this space-sellers have set "preferred-position" prices.

X. Position on Newspaper Page

In answer to the question, what is the best portion of a newspaper page for an advertisement, recommendations have been made as follows: the upper half, the right-hand portion, upper left-hand corner, upper right-hand corner. Or, in recognition of the need for considering the right and left-hand pages separately, the statement has been made that the upper outside corner is preferable.[1] These conclusions have been reached chiefly from laboratory investigations in which the subjects have been shown a "dummy" or a real newspaper or magazine containing advertisements or nonsense syllables distributed among the various portions of each page, and have been asked later to write down those items which they have remembered. The positions of those items mentioned the greatest numbers of times have been judged to be the most favorable. It should be noted that the conclusions regarding favorability of positions are for the most part based upon studies made with magazines. There is considerable doubt whether conclusions reached regarding positions in magazines are true also for newspapers.

Regardless of the advantageousness of these particular positions on a newspaper page it may be seen from even a cursory examination of the best type of newspaper that they are not equally open to the use of advertising matter. For there is a practice of newspaper make-up that crowds advertisements into certain portions of the page in order to reserve certain other portions for news. A circumstance which presents grave obstacles to the use of the positions which have been so laboriously designated as advantageous.

Specifically described, the practice of newspaper make-up, technically called "pyramiding," is to block up the advertisements first from the lower right-hand corner of the page. On some pages (usually the second page of the newspaper) only a small portion of this corner is used. But as an increasingly larger portion of the page (towards the middle and latter half of the paper) is consigned to advertisements the advertisements begin to run upward and leftward until often only a small portion of the upper left-hand corner is left for news. But this corner is usually thus reserved (unless the entire page is devoted to advertisements). Here is an apparent block to effective newspaper advertising, at least to that theoretically recommended by laboratory investigators.

---

[1] Adams, op. cit. p. 101.
Hotchkiss and Franken, op. cit. pp. 23-25.
D. Starch, Advertising. Scott Foresman and Co. 1914, Chicago, pp. 108-115.

In order to measure this block and express it in quantitative terms the writers undertook an investigation of actual journalistic practice. Choosing a newspaper that admittedly represents a high standard of journalistic make-up and advertising value, they tabulated the number of times the different quarters of pages were occupied by advertisements, and from this calculated the chances that an advertiser might secure a "favored" place for his advertisement.

The New York Times was used, and 1000 successive pages bearing advertisements and news were chosen from the issue of May, June and July, 1922. Those pages were omitted which contained nothing but advertisements, also those containing classified advertisements. Notation was made in each case as to whether or not the particular portion (half or quarter) contained any advertising matter, be it ever so small. Results are shown in Table I, and Figure 1.

TABLE I.

SHOWING THE NUMBER AND PER CENT OF UPPER AND LOWER, RIGHT AND LEFT PORTIONS OF LEFT AND RIGHT HAND PAGES OCCUPIED BY ADVERTISEMENTS IN THE NEW YORK TIMES DURING MAY, JUNE AND JULY, 1022.

505 Left Hand Pages								405 Right Hand Pages							
Lower Half				Upper Half				Lower Half				Upper Half			
Left Quarter		Right Quarter		Left Quarter		Right Quarter		Left Quarter		Right Quarter		Left Quarter		Right Quarter	
No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	%	'ON	%	'ON
175	35	300	77	150	30	370	73	115	23	335	08	115	23	205	00

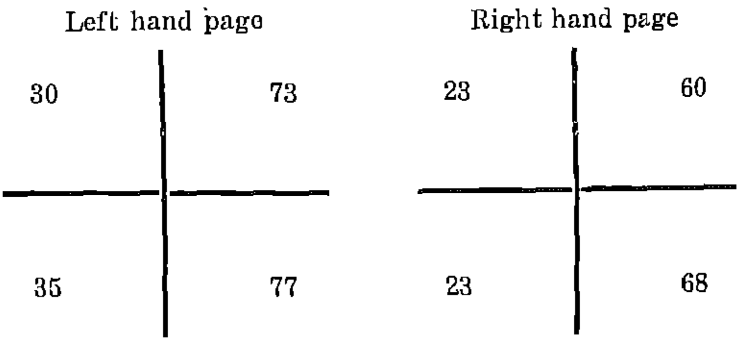


Fig. 1. Showing the per cent of pages on which the various quarters were occupied by advertising matter.

1. Left and Right-hand Pages.

Of the one thousand pages examined, 505 were on the left-hand page and 495 on the right-hand page. (It is interesting to note in this connection that *full-page* advertisements, which had been tabulated incidentally but not counted among the thousand shown in the table, appeared more frequently on the right-hand page. Of a total of 235 full-page advertisements, 120 or 53 per cent were on the right-hand page.)

2. Lower half.

a. Left-hand page. On the 505 left-hand pages the lower left quarter was occupied by advertisements in 175 or 35 per cent of the cases; the lower right quarter, in 390 or 77 per cent of the cases.

b. Right-hand page. On the 495 right-hand pages the lower left quarter was occupied by advertisements in 115 or 23 per cent of the cases; the lower right quarter, in 335 or 68 per cent of the cases.

3. Upper half.

a. Left-hand page. On the 505 left-hand pages the upper left quarter was occupied by advertisements in 115 or 23 per cent of the cases; the upper right quarter, in 295 or 60 per cent of the cases.

b. Right-hand page. On the 495 right-hand pages the upper left quarter was occupied by advertisements in 155 or 23 per cent of the cases; the upper right quarter, in 295 or 60 per cent of the cases.

CONCLUSIONS.

These results show that regardless of the alleged superiority of certain positions in newspapers, there are pronounced limitations to the likelihood that these positions may be secured. Thus although the right-hand page may be more desirable, it will be open for advertisers no more frequently than the left, since these results show it to be occupied only 495 times in 1000.

The upper outside quarter, which is recommended as being most valuable of all can be secured only 30 times in 100 on the left-hand page, though, fortunately it can be secured 60 times in 100 on the right-hand page. (For the sake of easier visualization, the frequency with which the various quarters of left and right-hand pages are used is shown diagramatically in Figure 1.)

The right-hand side of both pages is twice and even three times as likely to carry advertisements as the left-hand side. Which means that in the case of the left-hand page the majority of advertisements are going to be on the inside. And, unfortunately for the advertisers, this tendency is more marked in the case of the left-hand page than in the case of the right. These amounts may not hold good for all newspapers, especially for papers using the pyramid system more thoroughly than the New York Times. With such papers it is likely that the percentage of pages carrying advertisements in the upper left-hand corner would be still smaller.

From a journalistic point of view this quantitative investigation of the practice of pyramiding shows that the pyramiding progresses much more rapidly upward than leftward, the upper right portion of the page being occupied almost as frequently as the lower right portion. Though it should be remarked that the advertisements did not usually occupy the entire upper right quarter; tabulation was made if an advertisement merely extended into the area.

While this investigation furnishes no light concerning the relative values of the various positions on the newspaper page, it shows that there are distinct limitations to the ease with which certain positions can be secured, and it shows, at least for one high-grade newspaper, the size of the limitations in quantitative terms. It also suggests that the findings of laboratory studies regarding many other problems of advertising might profitably be checked up with actual conditions in order to show the practicability of their application in concrete situations.

AN EXPERIMENT IN TESTING ENGINE LATHE APTITUDE[1]

By Everett F. Patten
University of Wisconsin.

The investigation reported in this paper was carried out in co-operation with the Shop Practice Department of the College of Engineering, University of Wisconsin. It was able to profit materially by a preliminary investigation made along the same lines in 1921.[2] It was also favored by the fact that in the Shop Practice Course there were many students turning out supposedly identical pieces of work on fairly similar lathes. There was an added advantage in the fact that the men directing the work in the shop were university trained, and consequently were able to give such aid that the preliminary psychological analysis of the job was made more readily than would probably be the case in industry.

The particular type of work investigated was a course in elementary lathe practice. This involved four hours of shop work every week and two hours of instruction every two weeks. The students, following a combination of oral and printed instructions together with blueprints, turned out five pieces on an engine lathe during the semester. These were: a worm, a worm-wheel blank, a bevel-gear blank, a spiral-gear blank, and a planer bolt. The first four were turned out of cast iron and the last from rod steel.

The criterion of this investigation was the amount of skill developed in one semester of instruction. More specifically, the criterion was the quality of lathe work produced by the students during the course. Instructors' estimates are notoriously unreliable, as was found by the preliminary investigation and should never be used if an objective criterion can be obtained. Fortunately in this case the pieces themselves could be measured and a thoroughly objective criterion thus found in the accuracy with which these pieces correspond to the blueprint specifications. Working with great care Mr. Brecky of the Shop Practice Department took thirty-three micrometer measurements on the five pieces produced by each man measuring the various important dimensions of these pieces. Several thousand very accurate readings were thus obtained. In addition the general finish of each piece was scored on a percentage basis.

---

[1] From the Psychological Laboratory, University of Wisconsin. Grateful acknowledgment is hereby made to Prof. Clark L. Hull, Department of Psychology, and to Prof. W. L. Dabney, director of Machine Practice. Prof. Hull has supervised the investigation and given much helpful criticism, especially in the preparation of the manuscript. Prof. Dabney has given hearty co-operation and been of great assistance particularly in procuring subjects and in perfecting some of the tests.

[2] This investigation was reported in a master's thesis by D. W. Nelson. It is on file in the library, University of Wisconsin.

The measurements obtained were reduced to percentage grades. The deviation of each measurement from the dimension called for in the blueprint was first found. A check distribution was then made for each of the thirty-three measurements each distribution containing the deviations of all the men on that measurement. The median of each distribution was assigned the value of 85, the average grade given by the Shop Practice Department. The intervals of the distribution between the median and zero deviations were assigned values between 85 and 100 according to the size of the deviation. The fourth check from the opposite end of the distribution was assigned the value of 70 which is hardly passing on the school scale. The intervals of the distribution between this point and the median were assigned values between 70 and 85. The last three checks of the distribution were given below 70 or failure. Thus the larger the deviation from the standard the smaller would be the percentage grade. The mean of the grades for each individual piece was found and the average of the mean grades of the five pieces belonging to one man was computed in order to get his final grade in lathe working ability. Although this process took weeks of work the resulting criterion was very reliable. Indeed the large coefficient of total correlation (R) obtained later was probably due in no small degree to the care taken in obtaining the criterion.

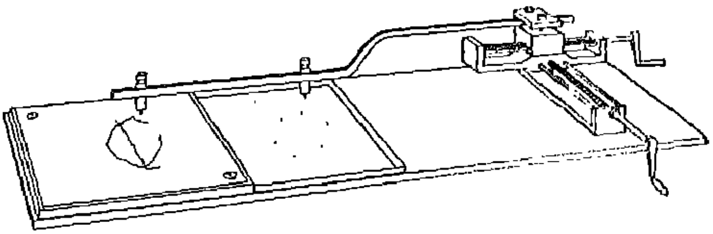
In proceeding to find tests which would measure lathe ability it was first found necessary to analyze that ability as far as possible. This was done by watching the men at work observing particularly wherein the good ones differed from the poor ones. As a result of this analysis eight tests were selected. Some of these at first glance would seem to have little connection with lathe ability in general, but they were included because they were thought to test a constituent factor. In selecting the tests an especial effort was made to get tests as dissimilar as possible so that there would be little duplication. The detailed analysis will be given with the discussion of the various tests.

(A) While studying the men at work it was noticed that the worker must be able to estimate two dimension space and to coordinate the motions of his two hands. He must also be quick and sure of movement. In order to test these abilities the "co-ordination machine" was devised.[1] The apparatus is shown in Fig.

---

[1] The need for such a machine was first seen by Dr. Hull, who worked out its general principles. The practical details of construction were worked out by Prof. Dabney. He also designed and made a number of improved models.

1. It may be best thought of as a compound *lathe* "rest" attached to a board. The action of the machine depended on two screws placed at right angles to each other. Turning these screws by means of small cranks caused a carriage to move in any desired direction in a horizontal plane. A long tracer arm was fastened firmly to the carriage. Projecting down from the middle of the arm was a small brass point. This touched the surface of a hard rubber plate which was fastened to the board near the main part of the machine. Embedded in the plate so that their tops were flush with the surface were six brass pegs arranged in a three-inch hexagon. These pegs were connected by an electric wire so that when the tracer point touched one of them a buzzer sounded. The object of the test was to start on signal from a certain point and make the circuit of the six brass spots as quickly as possible. The time taken to go around was measured by a stop-watch. A correlation of this with the criterion gave a -.423. This was the largest and most valuable correlation obtained.



Coordination Machine. Fig. I.

(B) Another measure was taken of this test, that of distance traveled. The tracer arm extended out beyond the contact point and had a pencil fastened to its extremity. This traced an exact record of the circular path on a piece of paper inserted for the purpose. A typical record is shown in Fig. 2. The distance traversed by this rather meandering line was measured by a specially devised instrument. The frame supporting the works of an ordinary watch was filed so as to expose the gear wheel on the shaft of the small second hand. The hands of the watch remained intact. The frame was so held in the hand of the measurer that the teeth of this gear could be run around the pencil tracing, following in detail all its turns. The distance traversed by this wheel was shown by the hands of the watch and could be read off at a glance. It will be seen that the circular distance was measured in units of revolutions of the small hand. One minute (or one revolution) was equal in length to 29.5 millimeters. The correlation of this measure with the criterion was +.192. This was rather surprising since it meant that the greater the distance, the better the grade on shop work. However, it was included in the final team of tests.

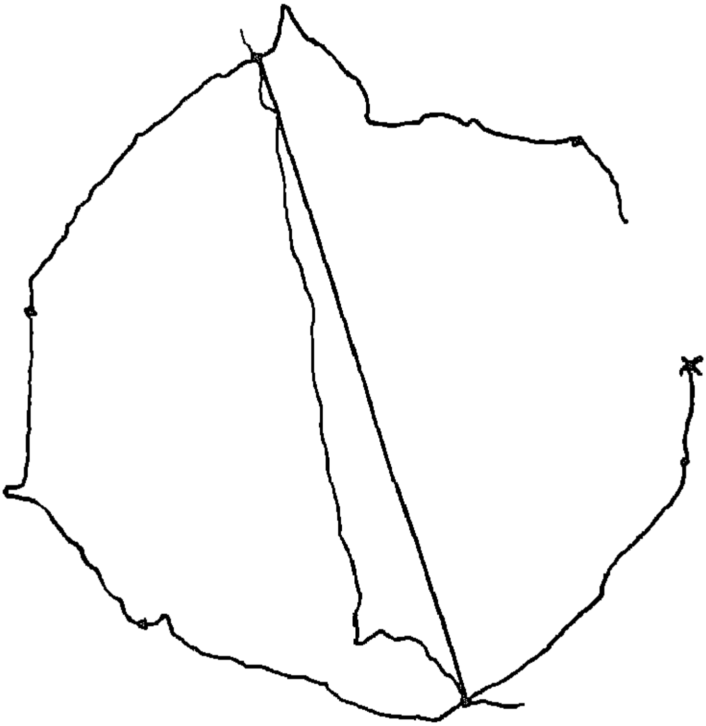


Fig. 2. Typical Tracing from Coordination Machine.

(C) A third score was also secured by the use of the co-ordination apparatus described above. It was observed that the worker must frequently cut a true surface in a diagonal direction using his eye to guide him. Accordingly the subject was required to run the tracer diagonally across between two of the most distant pegs going as quickly as possible. Here, as before, the time was taken with a stop-watch. The correlation of this measure with the criterion gave -.356. This was large but since it was thought that it would duplicate the first measure taken, that of time on the circular distance, measure C was discarded.

(D) The second measure of the diagonal co-ordination was made in deviations from a straight line between the two pegs. If a subject traced a straight line his record would coincide with the diagonal and there would be no deviation. However, if he wavered and made wrong moves, his record would deviate considerably from the straight line distance and would be poor in proportion to the amount of deviation. This deviation was measured by the area of surface included between the record tracing and the true diagonal as shown by a straight line drawn between the points. This was measured by means of a planimeter. The correlation of this measure and the criterion was -.034. This was so small as to be worthless. If the instructions had been to trace as straight a line as possible this might easily have proven a valuable test. It seems to have distinct possibilities.

It will be noted that the co-ordination test described above is a form of miniature test. It attempted as is usual in such cases to incorporate in a single test as many of the various elements necessary for success on the job and in as nearly the right proportions, as possible. Being a miniature it was portable and easily handled. One defect of the test is that it is undoubtedly influenced by training and thus is not a pure aptitude test.

(E & F) Further analysis of the job showed the necessity for close co-ordination between hand and eye as well as for quickness and accuracy of movement. Accordingly the number checking test was introduced. This was the standard Woodworth and Wells number checking test in which the subject was started on signal and told to cross out all groups of digits containing both 4 and 7. He was given one minute. The correlation of the number of right groups checked with the criterion was +.055. This was too low to be of use so was discarded from the final team of tests. The number of groups wrongly checked correlated -.035 so it was also discarded.

(G) It was believed that the possession of mechanical ingenuity would be important for success in lathe working so an effort was made to test for this quality. Accordingly an adaptation of Stenquist's lock assembly test was used. The lock was simple and inexpensive. It consisted of six parts which were: a rectangular lower case or plate, a dog, a spring, a bolt, a screw and an upper plate. The subject was shown a similar lock which was already assembled and told to put the parts of the other lock together like it. Although the lock seemed very simple, many had great difficulty in assembling it. Four failed entirely giving up after working from eleven to twenty-one minutes. The correlation with the criterion was -.199. Although not very high (-.33 was obtained last year) it had such a low correlation with the other tests that it was kept in the final team as tapping a real constituent of the particular aptitude under consideration.

(H) The ability to follow printed directions and blueprints is obviously a factor in learning lathe work. To measure this ability and to gain knowledge of the subject's carefulness and painstaking traits, a box-directions test was introduced. This test is the well-known Healy puzzle box, consisting of a small black cabinet containing a lock mechanism worked by three levers at the top and left sides of the box. The object of the test is to open this door. The necessary steps in the process were printed on a large card bearing an excellent diagram of the box. This was before the subject as he worked. This test proved even more difficult than the lock test. Ten men did not open the box at all, some giving up with the first failure. The correlation of the time with the criterion was -.270. This was large enough to warrant its use in the final team of tests.

(I) The number of wrong moves made was another measure taken on this test. However, this measure undoubtedly did not introduce any new element because errors caused the subject to start over again and thus increased his time. In addition the correlation was low (-.112) so this measure was eliminated.

(J) In doing good lathe work a great deal depends on the skill of the operator in taking accurate measurements. Accordingly the cylinder gauging test was introduced. A series of nine brass cylinders was very carefully made under Prof. Dabney's direction. These measured .970″, .965″, .960″, .940″, .930″, .900″, .885″, and .870″. It will be seen that there were three groups each composed of three cylinders. Each group or series was graduated by a different constant the same constant being used throughout a single group. The finest series changed by steps of .005″, the next by .010″, and the coarsest by .015″. The subject was given a six-inch machinist's rule, and a pair of calipers and told to find the cylinders which measure .965″ and .885″ as quickly as possible. The time taken to find the two cylinders constituted one measure. This correlated -.127 with the criterion and was too low to be of value.

(K) Another score on the cylinder test was the amount the chosen cylinder deviated in measurement from the one asked for. Thus if the subject chose the .930″ cylinder in place of the .965″ he was given a deviate score of .035. If he thought the .900″ cylinder measured .885″, he scored .015. The sum of these errors gave his deviation score, as .050. This measure correlated -.196 with the criterion. This seemed sufficiently high to be of use but it was finally decided to eliminate the test from the final team because it was really a trades test measuring acquired skill rather than aptitude.

(L) An effort was made to test ingenuity and reasoning ability by the match-squares test, though at the outset there seemed but a gambling chance of success. The match-squares test consisted of seventeen matches glued to a board in such a way as to make six squares in two banks of three each. The subject was shown these squares then given a mimeographed sheet on which short lines were drawn to represent the matches in the same position as on the board. He was told to cross out five of the short lines in such a way that if they were removed there would be left exactly three whole squares and no lines left over. The time was taken with a stop-watch. When through, the subject was asked whether he had ever done this or a similar puzzle before. The fact that many were already familiar with it together with the additional fact that twenty-two out of the sixty subjects could not do it at all seems to have destroyed what little diagnostic value the test may have had. The correlation with the criterion was only +.164. This was too small and also in the direction contrary to what would have been expected so the test was discarded.

(M) In watching the men at work it seemed probable that a keen sense of space relationship was necessary, particularly in guiding the cutting tool to its place and in finding centers. Accordingly the circle centering test was devised. A sheet of paper bearing a number of mimeographed circles, (see Fig. 3) was put before the subject. He was given a pencil and asked to make a dot in the center of each circle as quickly and accurately as possible. The task was complicated somewhat by the overlapping of some of the circles. The time required to place these dots was taken by a stop-watch and constituted one measure. This correlated -.201 with the criterion and since it was quickly and easily taken. Since the test was so cheap and portable it was kept in the final team of tests.

(N) Another measure was taken from this test the errors in locating the centers. These deviations from the true centers were measured translated into grades by a rather complicated process and these grades averaged to get the final grade for each man. Since this measure was not used the process will not be discussed further. The correlation with the criterion was fairly high (-.278) but was in a direction contrary to what was expected. Apparently the good subjects sacrificed accuracy to speed. Partly because of this, and partly because of the labor in scoring, this test was not included in the final team.

(O) It seemed reasonable to believe that if a man were skillful in handling woodworking tools he would also tend to be an adept in using other kinds, such as are employed in lathe work. Fol-

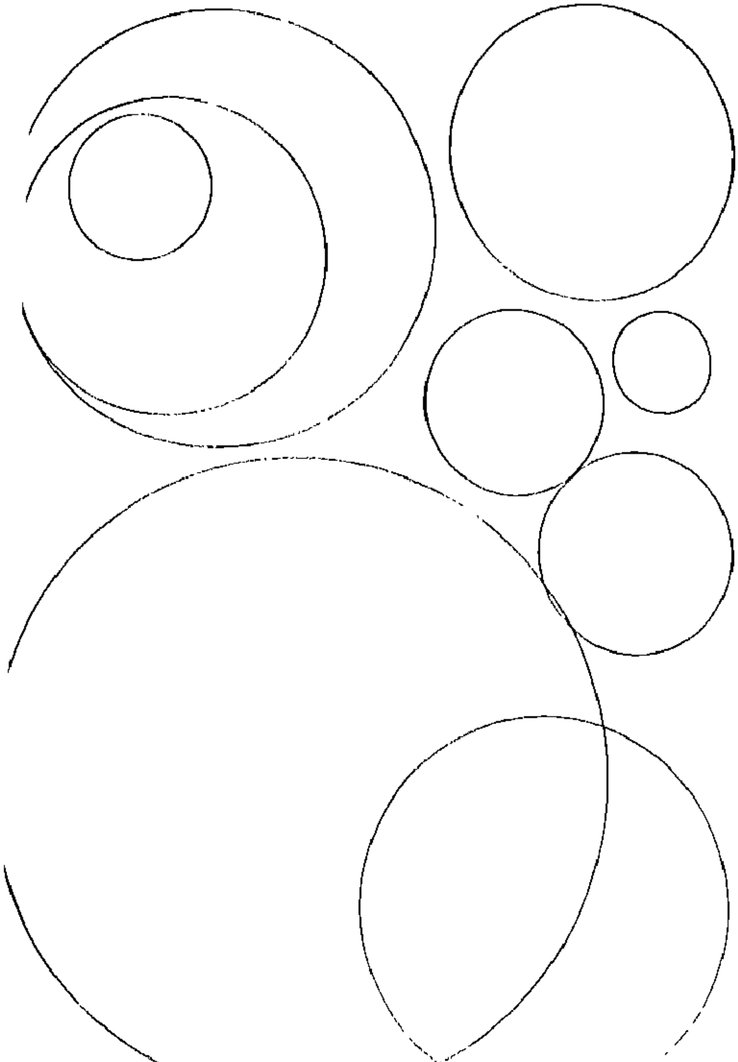


Fig. 3. Circle Centering Test.

lowing this line of reasoning the woodworking test[1] was given. The subject was asked to saw a three-inch square from a small board of definite size and shape, being given a saw, a square, a pencil and a special device for holding the wood such as is in common use in manual training shops. He was timed by means of a stop-watch. This measure had the surprisingly low correlation of –.119 with the criterion. It was discarded.

(P & Q) Two other measures were taken from the woodworking test one in deviation in dimensions from the three-inch square and the other in deviations from the true right angles. These latter measures were very difficult to make, requiring hours of almost microscopic work with a vernier rule, a square and a rule calibrated to one hundredths inch. The correlations in both cases were low, being –.095 and +.076 respectively. Both measures were discarded. It was a source of some surprise that the woodworking test did not have more diagnostic value.

The instructions for each test were given according to an unvarying formula so that the results would be as comparable as possible. The tests were given in a quiet room in a definite sequence. When the subject was admitted he went in regular succession from one test to the next, showing a card containing his name and number to each tester. By such devices a constant stream of subjects was kept flowing smoothly through the testing room. In order to obtain useful results, it was thought desirable to use a minimum of sixty records. Consequently over seventy men were tested so as to allow for the loss of ten through defective records from various causes.

A summary table of the correlations already given is as follows:

TABLE I

		r
*(A)	Criterion with co-ordination test, time	–.423
*(B)	Criterion with co-ordination test, distance	+.192
(C)	Criterion with co-ordination test, diagonal time	–.356
(D)	Criterion with co-ordination test, diagonal deviation	–.034
(E)	Criterion with number checking, rights	+.055
(F)	Criterion with number checkig, wrongs	–.035
*(G)	Criterion with lock test, time	–.199
*(H)	Criterion with box test, time	–.270
(I)	Criterion with box test, errors	–.112
(J)	Criterion with cylinder gauging, time	–.127
(K)	Criterion with cylinder gauging, deviations	–.196

---

[1] The woodworking test was devised by Mr. F. C. Mueller of the Psychology Department, University of Wisconsin.

TABLE I—cont.

(L) Criterion with match square, time	+.164
*(M) *Criterion with circle centering, time*	–.201
(N) Criterion with circle centering, grades	–.273
(O) Criterion with woodworking, time	–.119
(P) Criterion with woodworking, deviations in dimensions	–.095
(Q) Criterion with woodworking, deviations in angles	+.076

The variables finally used were renumbered for convenience:

Criterion =1
Co-ordination, distance =2
Co-ordination, time =3
Lock test, time =4
Box test, time =5
Circle centering, time =6

The means and standard deviations of these variables are shown in Table II.

TABLE II.

Means.	*Standard Deviations.*	*Unit of Measurement.*
M_1=84.046	$S.D._1$=3.689	Percentage grade on work pieces
M_2= 6.9406	$S.D._2$= .2653	Minutes-distance**
M_3= .9256	$S.D._3$= .3869	Minutes
M_4= 5.8585	$S.D._4$=5.089	Minutes
M_5= 4.78	$S.D._5$=2.627	Minutes
M_6= .3765	$S.D._6$= .2547	Minutes

The zero order correlations and intercorrelations are shown in Table III.

TABLE III.

r_{12}=+.192				
r_{13}= –.423	r_{23}= –.229			
r_{14}= –.199	r_{24}=+.142	r_{34}=+.056		
r_{15}= –.270	r_{25}=+.118	r_{35}=+.036	r_{45}=+.103	
r_{16}= –.201	r_{26}=+.022	r_{36}=+.170	r_{46}= –.032	r_{56}=+.059

The P. E. of each of the correlations in the first column is roughly .08.

It will be noted that although the correlations with the criterion are fairly high, the intercorrelations are exceptionally low. More than usual stress was laid on the latter in this investigation because it was thought that this principle is generally observed too little. To get the undiluted correlations with the influence of the other tests removed, the zero order correlations were subjected to partial correlation, using Yule's formula[4] and Rosanow's[5] general technique.

---

*Indicates the tests which were included in the final team. The above r's were all product-moment coefficients

**Distance in terms of circumference of gears on second hand.

$$r_{12.3}=\frac{r_{12}\ r_{13} \times r_{23}}{\sqrt{1-r_{13}^2}\cdot\sqrt{1-r_{23}^2}}$$

There were some fifty-two of these partials. Only those of highest order are given in Table IV:

TABLE IV

Partial correlations

$r_{12.3456} = +.101$

$r_{13.2456} = -.303$

$r_{14.2356} = -.211$

$r_{15.2346} = -.285$

$r_{16.2345} = -.100$

These partial correlations show but very little shrinkage and show that the experiment was unusually successful securing a set of tests highly uncorrelated with each other.

After the partial correlations were found, the coefficient of total correlation was computed. This correlation of the criterion with all the tests weighted in the best possible manner, was computed in two ways to check the arithmetic.[6] The formulae were

$$R=\sqrt{1-(1-r_{12}^2)(1-r_{13.2}^2)(1-r_{14.23}^2)(1-r_{15.234}^2)(1-r_{16.2345}^2)}$$

$$R=\sqrt{1-(1-r_{16}^2)(1-r_{15.6}^2)(1-r_{14.56}^2)(1-r_{13.456}^2)(1-r_{12.3456}^2)}$$

One correlation was .5549 and the other .5545 showing all the partials to be correct. This correlation is about as large as seems possible to secure from such highly selected groups as university students. It demonstrates the possibility of detecting aptitude in quality of lathe work produced.

In order to get the maximum yield from a team of tests as well as to use a set of test scores to predict probable success in a given vocation, a regression equation must be obtained. The general form[1] for this equation for six variables (criterion and five tests) is:

---

[4] Yule, Introduction to the Theory of Statistics. p. 239.
[5] Rosanow, Analysis of Mental Function, p. 30-43.
[6] Rosanow, op. cit., p. 40.

$$x_1 = r_{12.3456}\frac{SD_{1.23456}}{SD_{2.13456}}x_2 + r_{13.2456}\frac{SD_{1.23456}}{SD_{3.12456}}x_3$$

$$+ r_{14.2356}\frac{SD_{1.23456}}{SD_{4.12356}}x_4 + r_{15.2346}\frac{SD_{1.23456}}{SD_{5.12346}}x_5$$

$$+ r_{16.2345}\frac{SD_{1.23456}}{SD_{6.12345}}x_6$$

Where x_1 = the deviation of any subject from the mean of that variable (here the criterion), where x_2, x_3, x_4, x_5, and x_6 equal the deviation of any subject from the means of the respective variables, and where $S.D._{1.23456}$, etc., equal the partial standard deviations. These partial standard deviations are obtained by a formula of the type:[2]

$$S.D._{1.23456} = S.D._1 \sqrt{1-r^2_{12}} \sqrt{1-r^2_{13.2}} \sqrt{1-r^2_{14.23}}$$

$$\sqrt{1-r^2_{15.234}} \sqrt{1-r^2_{16.2345}}$$

The partial standard deviations are shown in Table V:

TABLE V.

$S.D._{1.23456} = 3.070$ $S.D._{4.12356} = 4.881$

$S.D._{2.13456} = .248$ $S.D._{5.12346} = 2.483$

$S.D._{3.12456} = .343$ $S.D._{6.12345} = .247$

Substituting these, together with the partial correlation coefficient, in the formula for the regression equation and solving we have:

$$x_1 = 2.397x_2 - 3.252x_3 - .132x_4 - .353x_5 - 1.996x_6$$

This, however, is not a very convenient form of the regression equation for practical use. Substitutions were accordingly made for the small x's so that instead of deviations from means were had individual scores minus the mean of that variable (X_n-M_n. The formula then becomes:

---

[1] Yule, op. cit., p. 241.

[2] Yule, op. cit., p. 237.

$$X_1-M_1=2.397(X_2-M_2)-3.252(X_3-M_3)-.132(X_4-M_4)-.353(X_5-M_5)$$
$$-1.996(X_6-M_6)$$

Substituting values for the various means (found in Table III) and solving, the following equation was secured:

$$X_1=73.635+2.397X_2-3.252X_3-.132X_4-.353X_5-1.997X_6$$

It will be noted that the values X_2 to X_6 represent gross scores of any given subject on the present team of tests while X_1 represents his probable grade in a course such as previously described. Now if a person of unknown aptitude in this kind of work should be given the above tests, his test scores might be substituted in the above equation and his most probable grade in such a course be predicted with some precision. To take an actual case, subject number 9 made in this team of tests, the following scores:

Co-ordination, distance=6.6 revolutions
Co-ordination, time=1.05 minutes
Lock test=1.17 minutes
Box test=3.37 minutes
Circle centering, time=.35 minutes

Substituting these in the equation, the result is:

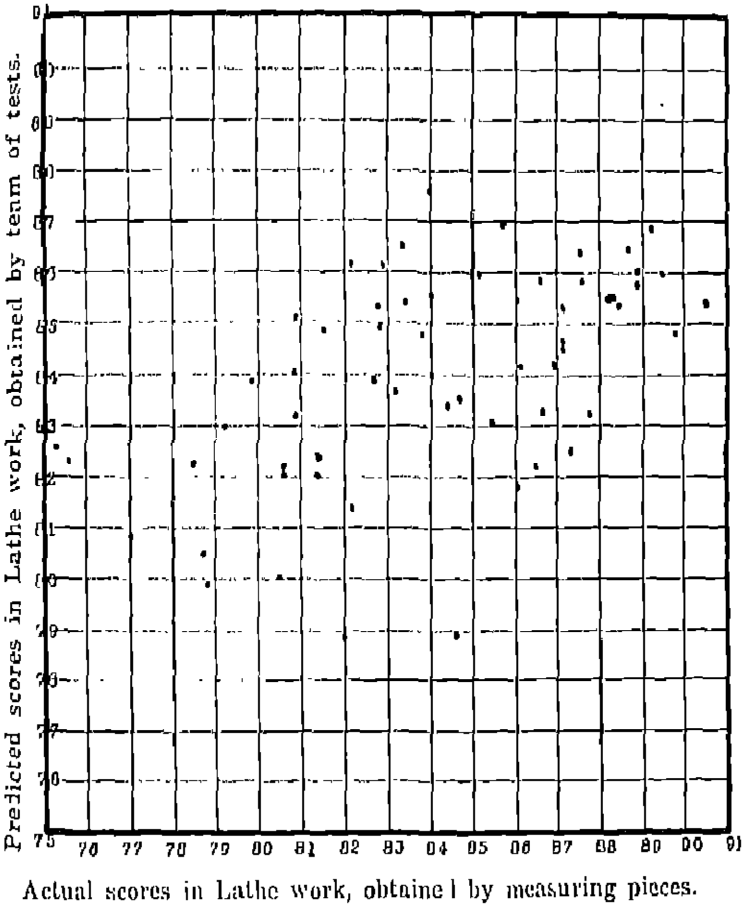
$$X_1=73.635=2.397(6.6)-3.252(1.05)-.132(1.17)-1.996(.35)$$

Solving, X_1=84.008. This man actually scored 84 in his five pieces.

To lighten the work of prediction, both for this investigation and for the one to be carried on with a new group of subjects here next semester, a chart was prepared containing all the products of the probable test scores and their respective weights. The scores were then predicted for all the sixty subjects, both to confirm the R and to observe the detailed relation of individual predicted scores to the actual scores. This is shown in detail in the scatter diagram, Fig. 4.

FIGURE 4.

Scatter Diagram.



RACIAL DIFFERENCES AS MEASURED BY THE DOWNEY WILL-TEMPERAMENT TEST

By JOHN H. McFADDEN and J. F. DASHIELL

INTRODUCTION

Study of character and personality traits is, by its very nature, a perplexing problem, since there are a multiplicity of factors entering into the make-up of any given individual, and since the organization of these factors varies so greatly with each individual. Several tentative classifications of the traits and of the factors governing the combinations of the traits have been made. Woodworth[1] analyzes personality into four factors: Physique, Chemique or Temperament, Instincts, and Intelligence. F. H. and G. W. Allport[2] divide tendencies into: Intelligence; Temperament, including Emotional Breadth and Emotional Strength; Self-expansion, including Extro-Introversion, Ascendance-Submission, Expansion-Reclusion, Compensation, and Insight and Self-Evaluation; and Sociality, including Social Participation, Self-seeking and Aggressive Self-seeking, and Susceptibility to Social Stimuli. Tests for these were given, and the results checked by estimates of character and personality given by friends of the subjects tested. The above division of tendencies is considered by the writers as merely a tentative one which may be of some use in guiding further character analysis. Watson[3] gives a form for analysis of personality, and Wells[4] classifies the healthy mental reactions as distinguished from unhealthy ones. H. T. Moore and A. R. Gilliland[5] studied aggressiveness as a personality trait, isolating it as far as possible by selecting two groups of students, one composed of students rated as the most aggressive in their class, and the other composed of students rated as the least aggressive in their class. The following tests are suggested as measurements of aggressiveness: eye movements; distraction from staring, from electrical shock, and from visual presentation of a snake; and a word association test. In the above tests, the first group scored high, and the second group scored less, thus indicating the value of the methods used. R. O. Filter[6] studied Self-Assurance and Speed of Decision, concluding that one test was not sufficient for each trait, but that two or more should be given. Also, he concludes that there was not a very close correlation between speed of decision and self-assurance. The test used in the present study of racial differences is the Downey Will-Temperament Test[7] *formulated by Dr. J. E. Downey, of the University of Wyoming.*

---

[1] Woodworth, R. S. *Psychology, a Study of Mental Life.* Henry Holt and Co., New York, 1921.

[2] Allport, F. H. and Allport, G. W. *Personality Traits: Their Classification and Measurement.* Journal of Abnormal Psychology and Social Psychology, Vol. 16, No. 3. April, 1921.

[3] Watson, J. B. *Psychology from the Standpoint of a Behaviorist.* J. B. Lippincott Co., Philadelphia.

[4] Wells, F. L. *The Systematic Observation of the Personality—In Its Relation to the Hygiene of the Mind.* Psychological Review, Vol. 21, 1914, pp 205-332.

[5] Moore, H. T. and Gilliland, A. R. *The Measurement of Aggressiveness.* Journal of Applied Psychology. Vol. 3, No. 2, June, 1921.

The particular problem here studied is the comparison of temperament as found in individuals of the white and negro races. Previous studies have been made regarding the comparative intelligence of the two races. Mayo[8] finds that, while the "data points clearly to a measurable degree of mental difference," the differences are of degree, and not of kind, and that mental variability is slightly less in negro high school students than in white high school students. Ferguson[9], using the Woodworth and Wells Mixed Relations Tests I and II, a form of the Ebbinghaus Completion Test, a Cancellation Test, and one of the Columbia Maze Tests, finds the intellectual performance of the general colored population to be approximately seventy-five per cent as efficient as that of whites, with the possibility of negroes more nearly approaching whites in reasoning based upon concrete objects than in reasoning based upon mental representations of the objects. He finds a practical equality in sense capacity in the perceptive and discriminative abilities, and, since the difference must exist in the central elaborative powers upon which thought more directly depends, he infers that such neural differences as may be found will probably be mainly in the constitution of the cortical neurons, rather than elsewhere in the nervous system. This point will be taken up later. Ferguson also states that, while no certain conclusion as to racial variability can be reached, it would be reasonable to suppose that the negroes would vary more than whites, since the so-called negro race is not nearly as homogeneous as the white race. Derrick[10] finds that the negro college student is older than the white, that their average intelligence quotient is nine points lower than the average white intelligence quotient, and that the white college student is about seventy-seven per cent as variable as the negro college student.

---

[6] Filter, R. O. *An Experimental Study of Character Traits.* Journal of Applied Psychology, Vol. 5 No. 4, December, 1921.

[7] Downey, J. E. *The Will-Profile. A Tentative Scale for Measurement of the Volitional Pattern.* Univ. of Wyoming Bulletin, Vol. 15, No. 6A, Feb., 1919.

The Adolescent Will-Profile. Journal of Educational Psychology. Vol. 11, No. 3, March, 1920.

Some Volitional Patterns Revealed by the Will-Profile. Journal of Experimental Psychology, Vol. 3, No. 4, August, 1920.

Manual of Directions for Testing. World Book Co., Yonkers-on-Hudson, N. Y., 1921.

[8] Mayo, M. J. *The Mental Capacity of the Negro.* Archives of of Psychology Monograph No. 28, 1913.

[9] Ferguson, G. O., Jr. *The Psychology of the Negro.* Archives of Psychology, Monograph No. 36, April, 1916.

Turning from previous work on intelligence, and considering other studies of racial differences, we find several reports on different traits. Stetson[11] finds practically no difference in memory in the two races, though there may be some doubt as to the value of this contribution. Bache[12] finds the negro quicker in his reaction time than the white man, and Garth[13] finds that there are slight differences in their work curves—that there is a tendency of the negros to do more of their work in the first half of the allotted time than for the white workers. This would indicate that the negroes are more quickly aroused to work, but that, after the first spurt, their effort diminishes. Mitchell, Rosanoff and Rosanoff[14] in a study of association, find that negro children, on the whole, show further departure than white ones from the normal adult association standard.

Before going on into our particular problem, it is necessary to note some cautions which experimental work has given. Pressey[15] warns that grade norms may vary very greatly in various schools, that there may be different age distribution, and that other factors may enter in. And, in a comparison of students of different races, it is all the more important to bear this in mind, since grade norms will inevitably vary much more greatly in schools for different races than in different schools for students of the same race. Arlitt[16] finds a tendency for greater variation (in results from intelligence tests) in the same race but different social status than in the same social status but in different races. This latter point will be considered in the next section. Further, as Gatewood[17] points out, there is a tendency to rate negro children above their mental status, as compared with the tendency to rate whites at or below their true mental status.

---

[10] Derrick, S. M. *A Study of the Intelligence of White and Colored College Students.* Journal of Applied Psychology. Vol. 4, pp. 316-329, December, 1920.

[11] Stetson, C. R. *Some Memory Tests of Whites and Blacks.* Psychological Review, Vol. 4, pp. 285-289, 1897.

[12] Bache, R. M. *Reaction Times With Reference to Race.* Psychological Review, Vol. 2, pp. 475-480, 1895.

[13] Garth, T. R. *White, Indian and Negro Work Curves.* Journal of Applied Psychology, Vol. 5, No. 1, March, 1921.

[14] Mitchell, I., Rosanoff, I. R., and Rosanoff, A. J. *A Study of Association in Negro Children.* Psychological Review, Vol. 26, September, 1919.

[15] Pressey, S. L. *The Problem of the Unselected Group in the Standardization of Tests.* Journal of Applied Psychology, Vol. 5, No. 1, March, 1921.

[16] Arlitt, A. H. *On the Need for Caution in Establishing Race Norms.* Journal of Applied Psychology, Vol. 5, No. 1, March, 1921.

[17] Gatewood, E. L. *Teachers' Estimates of Negroes and Whites.* School and Society, Vol. 9, pp. 90-91, 1919.

SUBJECTS AND TESTS

The subjects tested were white and negro high school and college students. It was hoped, by choosing such groups, to eliminate many of the factors of selection, and to make the groups more directly comparable, and also to fix a general level of intelligence between the groups. It was found, however, that, in general, negro students of each grade were slightly older than white students of the same grade. This is constantly found in comparisons of white and negro students. Graphs and tables follow, which endeavor to analyze the general results, and to interpret them in terms of equal ages and equal grades. These results are based upon the records of 38 high school and 39 college students from each race.

The thirty-eight white high school students were tested at the Chapel Hill High School, and were chosen at random, nearly all of the male students being taken. A note should be made here concerning the make-up of this high school, since it is a little out of the ordinary. The high school students are drawn from three fairly distinct social classes: those from the surrounding farm and mill districts, those whose parents are in business in the town of Chapel Hill, and those whose parents are members of the Faculty of the University of North Carolina. A survey of general intelligence of these students with regard to their respective social classes was made in 1917 by Chase and Carpenter and again in 1921 by Dashiell and Glenn, who found that the social levels and intelligence levels were strongly related. As might be expected, the Faculty children led in general intelligence, followed by the Town children, with the Out-of-Town group last. Chapel Hill High School embraces the 8th, 9th, 10th, and 11th grades, with seven years of grammar school work preparatory.

The thirty-nine white college students were taken from classes in psychology at the University, the entire class in Abnormal Psychology being tested, and the balance volunteering from introductory classes. Except for the very few freshmen and seniors, it is a fairly typical group of college students.

For the negro subjects, students in three negro institutions were tested. The National Training School, at Durham, N. C., furnished five college students and all except two of the high school students. There is some doubt as to the validity of all the results from this group, since the testing was stopped, after forty subjects had been secured, upon evidence that some of the students were being coached by their fellows. This would naturally affect the Reaction to Contradiction test, and it seemed to affect the Motor Inhibition and Co-ordination of Impulses tests, since there was an unnatural tendency for the subjects to write excessively slowly in the preliminary warming-up (Test II, 3), and, for the Co-ordination of Impulses, to take great pains in adhering to the space limitations, but to fail in the time limitations. These tendencies were not noted with other groups, either of negroes or of whites. The doubt concerning coaching, and the fact that these tests and the norms for grading them were developed primarily for adults, would render the high school comparisons somewhat inconclusive, and would emphasize the results from the college groups.

The second group of negro students was from Shaw University, in Raleigh, N. C. The men tested were selected by the President of that University. Shaw University is one of the best negro institutions in the South in regard to college ranking. Eighteen men were tested there, during a period of three days. The writer does not believe that any of the men in this group were coached.

The last group of negro students was from the Negro Agricultural and Technical School, at Greensboro, N. C. This is a state institution, but, owing to the fact that agricultural schools usually have lower standards of admission than the Colleges of Liberal Arts in the various Universities, it is possible that a comparison of the men in this group with men from the University would be unfair. Still, the three negro schools visited are typical of the institutions of that race, and the men afford a better basis of comparison, being more widely distributed, than if all the subjects were taken from one institution. At the Agricultural and Technical School, eighteen men were tested during a period of three days, sixteen of the men being of college grade, and two of high school grade.

Although precautions were taken to standardize the subjects in regard to education, it must be admitted that the criticism of Arlitt[15] regarding differences in social status holds good here. Derrick[16] reports that "White boys from average environment go to college nearly twice as frequently as negro boys from average negro environment"; thus, the negro college students are more highly selected from their environment than the white students from theirs. The group of negro agricultural college students furnished a partial check on this condition, but for the best study in race comparison, it would be necessary to select a large number of individuals from all environments. The present comparison can claim to hold good only for college and high school students of the two races.

---

[15] Arlitt, A. H. *On the Need for Caution in Establishing Race Norms.* Journal of Applied Psychology, Vol. 5, No. 1, March, 1921.

[16] Derrick, S. M. *A Study of the Intelligence of White and Colored College Students.* Journal of Applied Psychology, Vol. 4, pp. 316-329, December, 1920.

To better insure uniformity in results, all of the tests here reported were given by one of the writers and were graded by him. The grades were checked by the same writer and by a student trained in the method of grading.

It is not thought necessary to give an exhaustive description of the Will-Temperament Test, but definitions of the traits scored, and a statement of the value of the scores in each, are necessary for a full understanding of the results. In the order in which they are listed by Prof. Downey[20], they are:

Speed of Movement: Speed of movement of person relative to size and age. A score of "10" here indicates a very high speed, grading down to "1", which indicates very slow movement.

Freedom from Load or Inertia: Tendency to work at one's highest speed without external pressure, little tendency to relax speed, quickness in warming up to a task. A score of "10" indicates practically uniform speed, "1" indicates great increase, which may mean that the subject usually works more slowly than he could—that he has excessive "load" or weight of inertia.

Flexibility: Ease and success in readjustment; capacity to modify one's routine reactions. A score of "10" here indicates highest success, "1," practically failure.

Speed of Decision: Quickness in reaching a decision or conclusion. A slow reaction here may be due to caution or conservatism in weighing the elements involved in a situation, or be caused by one's being side-tracked by irrelevant matters or a rambling procedure. Here "10" indicates a speedy decision, "1", a very slow decision.

Motor Impulsion: This trait refers both to impetuosity and energy of reaction. Consider the ease with which the brakes or inhibitions can be removed, and also the tendency to explosive reaction when the brakes are actually off. A "10" here indicates explosive or expansive reaction under distraction, grading down to a "1," which indicates inhibition or withdrawal into one's self.

Reaction to Contradiction: This refers to the degree of confidence with which one maintains his opinion against contradiction. A "9" or "10" reaction signifies an aggressive subject—the burden of proof is thrown on the person who does the contradicting; "7" and "8" are confident reactions, but reasons are cited for one's confidence and the burden of proof is accepted; "5" and "6" are tactful reactions; below "5" there is a grading down to complete failure to assert one's own opinion.

Resistance to Opposition: The vigor with which one reacts immediately to a blocking of one's purpose. It grades from "10", a strenuous reaction, to "1", complete passivity in the face of opposition.

---

[20] Downey, J. E. *Manual of Directions for Testing.* World Book Co., Yonkers-on-Hudson, N. Y., 1921.

Finality of Judgment: Confidence in previous decision, willingness or unwillingness to reopen a question once closed. A "10" indicates refusal to reconsider, or a very rapid reconsideration, a "1" indicates extremely slow reconsideration. Sometimes the subject spends as much time on review as he does on the original decision. The score in this trait should be interpreted while bearing in mind the score or time spent on the original decision.

Motor Inhibition: Capacity to keep in mind a set purpose and achieve it *slowly*. It involves power of motor control, imperturbability, and patience. A score "10" indicates highest success, a "1" indicates practically no inhibition.

Interest in Detail: Attention to details. This trait is not equivalent to accuracy, but rather it indicates thoroughness, which may or may not be accompanied by analytical thought or insight. A "10" indicates highest care for details, a "1" shows very little care.

Co-ordination of Impulses: Capacity to execute a double task without preliminary trial, capacity to handle a complex situation successfully without forgetting either factor involved. This trait is probably allied to keeping one's head in a confusing situation. A "10" here indicates greatest success, a "1" indicates practically failure.

Volitional Perseveration: Absorption in a task; willingness to keep "plugging away" at it because the examinee sets up a goal for himself. A "10" here is also the most desirable score, while a "1" indicates little desire to take pains in perfecting a piece of work that the individual takes personal pride in.

The scores in the Downey tests were obtained by taking the records of a large number of individuals, for some given trait, and assigning score values in such a way that 10 per cent of the persons tested attained each score. Thus, in a given group, we could hardly expect to find normal distribution curves for the various traits, but rather a tendency for an equal number to attain each score. Further, the total score made on the test is not as significant as the grouping of the scores in the various traits, although it has been considered that a high total score indicates a "strong" personality, and a low total score indicates a "weak" personality. Also, the tests and grading norms were devised for adults, and this should be taken into consideration when these tests are given to younger subjects.

G. M. Rush[21] gives a criticism of the Will-Temperament

---

[21] Rush, G. M. *A Preliminary Study of the Correlations Between Estimate of Volitional Traits and the Results from the Downey "Will-Profile."* Journal of Applied Psychology, Vol. 5, No. 2, June, 1921.

Test, in regard to correlation of the results and estimates. He finds, for fifteen subjects, very little correlation between the estimates of the traits and the scores in those traits as tested by the Will-Temperament Test. Prof. Downey[33]; however, using a slightly different method of estimates, found sufficient correlation to justify the tests.

RESULTS AND INTERPRETATIONS

Table I contains the summary of means and medians of the scores in the various traits for the two races, divided into college, high school, and combined groups, with the standard deviation from the mean in the various traits for the two college groups.

TABLE I.

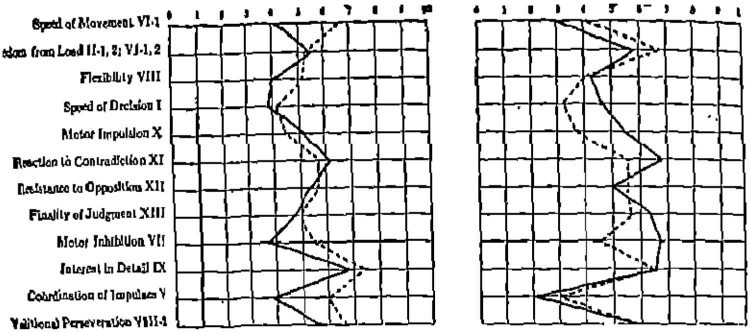
	College Students			High School		Combined	
	WHITE STUDENTS						
Trait	Mean	S. D.	Median	Mean	Median	Mean	Median
Speed of Movement	6.77	2.531	7	4.13	3	5.46	5
Freedom from Load	5.41	2.527	5	6.71	7.5	6.06	6
Flexibility	5.15	1.740	5	4.13	4	4.65	5
Speed of Decision	4.10	2.888	4	3.21	3	3.66	3
Motor Impulsion	4.51	1.781	5	3.76	4	4.14	4
Reaction to Contradiction	5.79	2.451	7	5.50	6	5.65	6
Resistance to Opposition	5.77	2.731	7	5.50	7	5.63	7
Finality of Judgment	4.92	2.392	5	5.60	5	5.25	5
Motor Inhibition	5.41	3.409	6	4.60	5	5.01	5
Interest in Detail	7.33	1.489	7	6.44	6	6.89	7
Co-ordination of Impulses	6.02	2.882	7	3.05	2.5	4.82	4
Volitional Perseveration	6.64	2.956	8	5.81	6	6.23	7
	NEGRO STUDENTS						
Trait	Mean	S. D.	Median	Mean	Median	Mean	Median
Speed of Movement	4.10	3.070	3	3.08	2	3.60	3
Freedom from Load	5.48	2.716	5	5.87	5	5.67	5
Flexibility	4.00	2.000	4	4.23	4	4.11	4
Speed of Decision	3.87	3.048	2	4.71	4	4.28	3
Motor Impulsion	5.15	1.955	5	5.47	5	5.31	5
Reaction to Contradiction	6.10	1.997	7	6.81	7	6.45	7
Resistance to Opposition	5.33	2.641	5	5.05	5	5.19	5
Finality of Judgment	4.77	2.301	5	6.29	7	5.52	5
Motor Inhibition	3.84	3.519	3	6.76	8	5.28	6
Interest in Detail	6.82	2.146	7	6.52	7	6.67	7
Co-ordination of Impulses	3.87	2.838	3	2.13	1	3.01	2
Volitional Perseveration	5.69	2.524	6	5.80	7	5.74	7

---

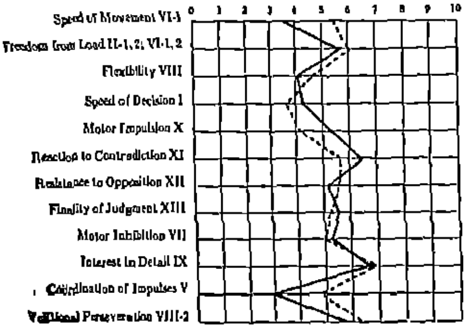
[33] Downey, J. E. *The Will-Profile, A Tentative Scale for Measurement of the Volitional Pattern.* Univ. of Wyoming Bulletin. Vol. 15, No. 6A, Feb. 1910.

Attention must be called to the fact that a composite profile plotted from mean or median scores, as given in Figure 1 and Figure 2, must be read with caution, since there is so great possibility of individual variability in the scores in each trait. Also, these profiles do not indicate, more than in a very general way, tendencies toward racial patterns. However, from Table I, and Figures 1 and 2, we may say in general that the "average" white student has greater speed of movement, that he has slightly greater freedom from load—that is, he tends to work on the weighing of the evidence, or less decisiveness; that he is slightly average more nearly at his greatest speed; that he has slightly greater flexibility and ability to meet new conditions in life; that he is slightly slower in making decisions—which may mean a better less liable to become explosive under distraction from conflicting stimuli; that he is less forceful when contradicted, which may mean

FIGURE 1. Mean Scores.



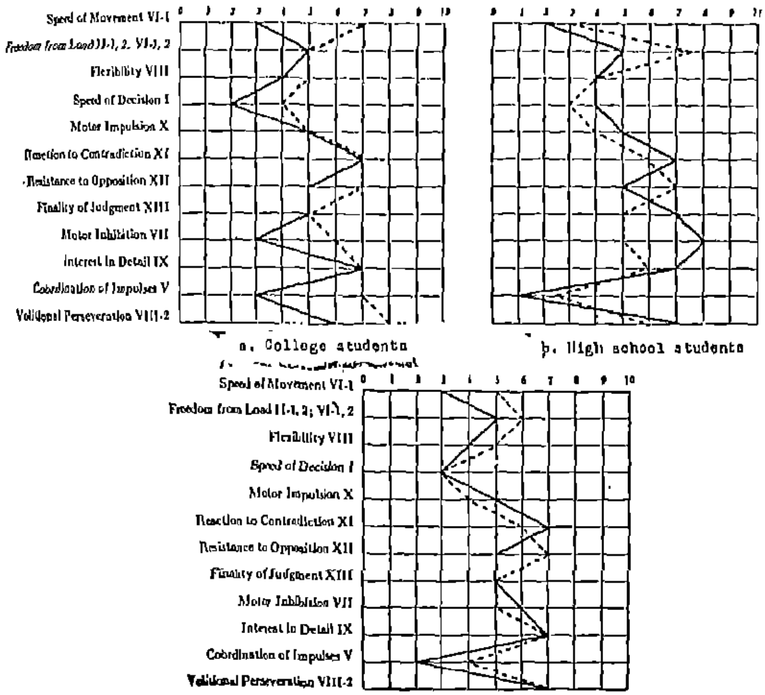
a. College students b. High school students



c. Combined

a greater degree of social adjustment (a score of "5" is the tactful response), or a greater degree of suggestibility; that he is firmer when confronted by obstacles or physical opposition; that he *takes about the same time, or slightly longer in reconsidering*

FIGURE 2. Median Scores.



a. College students

b. High school students

c. Combined

his decisions, which may be due to greater care, or to greater *indecision; that he has, as a college student, greater ability or* desire to restrain his impulses, and, as a high school student, less ability or desire (but here must be remembered the criticism given above, that there was some evidence of coaching upon the results from negro high school students); *that he has practically the same* interest in detail; that he has a markedly greater ability to coördinate his impulses, or to meet confusing situation; that he has a slightly greater tendency to volitional perseveration.

Figure 3 shows the distribution of the individual scores in each trait, for the high school students, the college students, and the combined groups, of both races. Score values are indicated on the horizontal axis, number of individuals on the vertical.

The median profile of the high school students (Figure 2b) shows emphasis on the traits which combine to suggest an aggressive personality, a result which was found by Dr. Downey[23] in testing high school freshmen.

A point may here be brought out in relation to the negro's tendency to resist contradiction more strongly than the white. Dr. Bevis[24], tabulating the frequencies of various forms of psychoses among the Southern negroes, reports that dementia praecox stands at the head of the list of their psychoses, with the catatonic form occurring about twice as often as in the white. Since one of the characteristic symptoms of the catatonic form of dementia preacox is negativism, the result of this test checks fairly with Dr. Bevis' report. Also, (and cf. No. 6 in "Summary") we might expect a greater degree of aggressiveness from the negro when he is allowed to reveal his inner self. As a race, they are generally oppressed or held down, a fact which might lead us to look for a reaction in the direction to greater aggressiveness when unchecked. Observations of Northern negroes have been reported, where the negro, freed from race restraints, has swung to the other extreme and has been rather too aggressive; and the tendency of the negro drivers to beat their mules more mercilessly than do white drivers, is a matter of observation. This is apparently a case of "inferiority complex" and "compensation".

The lower score of the white students in motor impulsion might have been expected. Odum[25] says, "The two general characteristics, then, the negro's emotional nature are the lack of restraint and the consequent extreme expression of the feelings." These would naturally be allied with greater motor impulsion.

Various investigators have made different reports regarding the variability of the negro as compared with the white. Ferguson[26] says that no certain conclusion as to racial variability can be reached, adding "It may finally be remarked with regard to the relative variability of whites and negroes, that it would not be at all surprising if groups of so-called negroes were definitely shown to be more variable than comparable groups of whites. For the groups that are generally called negroes are composed of in-

---

[23] Downey, J. E. *The Adolescent Will-Profile.* Journal of Educational Psychology, Vol. 11, No. 3, March, 1920.

[24] Bevis, W. M. *Psychological Traits of the Southern Negro.* American Journal of Psychiatry, Vol. 11, No. 1, July, 1921.

[25] Odum, H. W. *Social and Mental Traits of the Negro.* Columbia University Studies in History, Economics and Public Law, Vol. 37, No. 3, Whole No. 99, 1910.

[26] Ferguson, G. O., Jr. *The Psychology of the Negro.* Archives of Psychology, Monograph No. 36, April, 1916.

FIGURE 3.

Distribution of Individual Scores in the Different Traits.

————Negroes - - - -. Whites

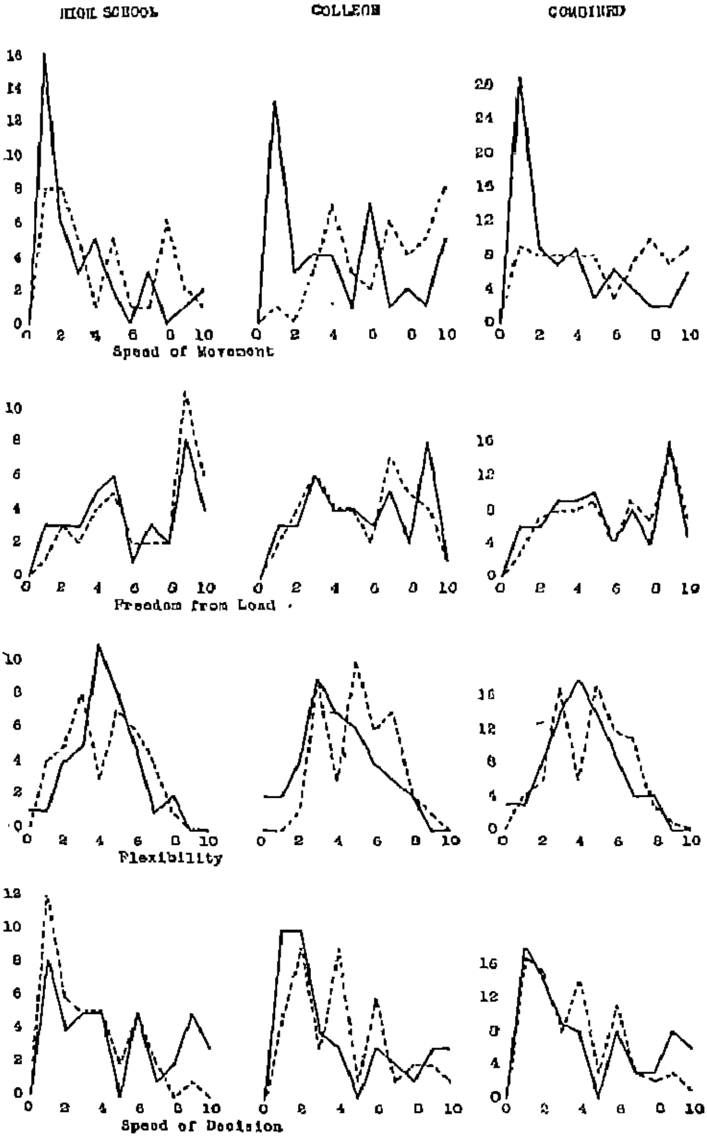
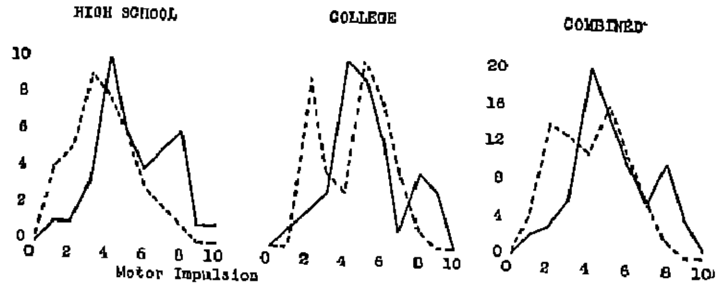
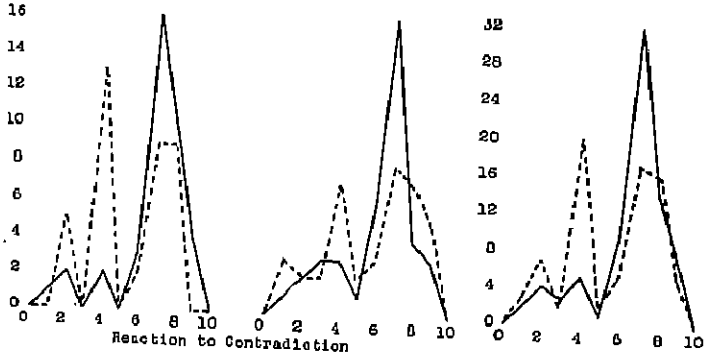


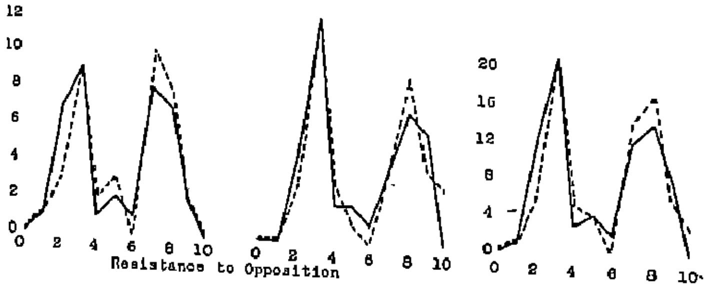
FIGURE 3. (cont.)



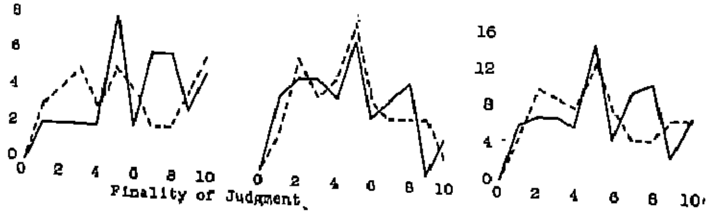
Motor Impulsion



Reaction to Contradiction

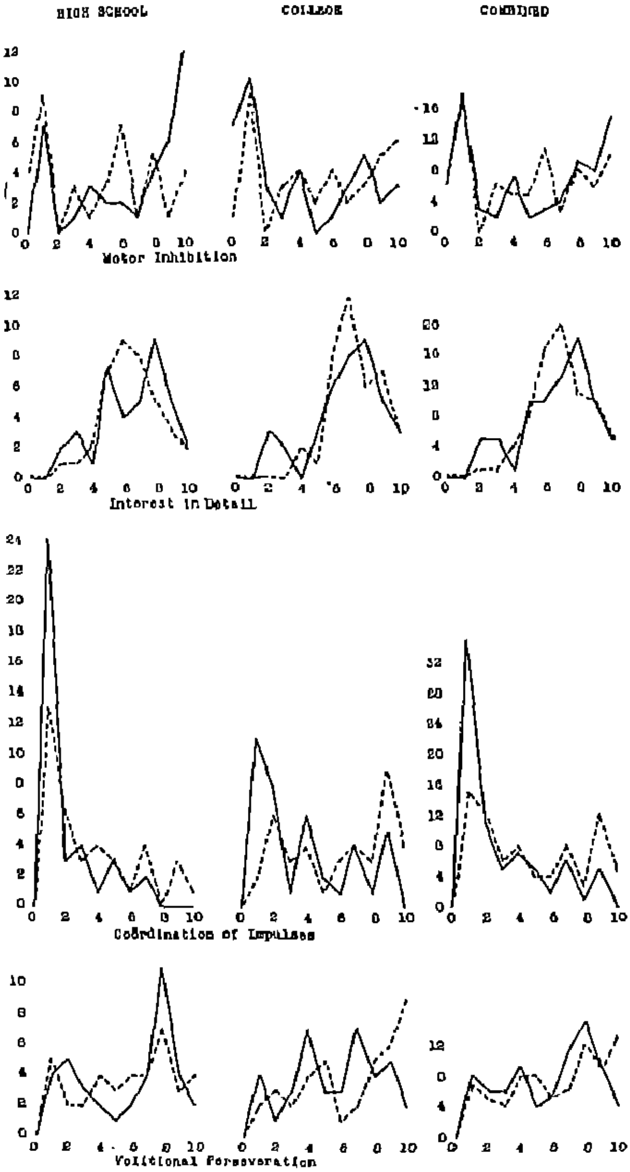


Resistance to Opposition



Finality of Judgment

FIGURE 3. (cont.)



dividuals ranging from pure negroes to persons almost white, and it would be reasonable to suppose that such groups would vary more than would homogeneous white groups." Mayo[27] states that mental variability is less in negro high school students than in white high school students, the ratio being 93 to 100. Derrick[28] finds the opposite in regard to college students, when compared in respect to intelligence. He says "The white college student was about 77 per cent as variable as the negro." Nor is anything definite concerning variability found in the results from the Will-Temperament Tests. Inspection of the distribution curves for the scores in the various tests (Figure 3) shows practically no preponderance of either race. Comparison of the standard deviations from the mean in the scores made by college students, Table II, shows that the negro deviations exceed those of the whites in seven of the traits, and that the amount of excess is relatively greater. There would, then, seem to be a slight but not striking tendency for the negro to be more variable.

TABLE II.

Comparison of the Standard Deviations of the Scores Made by the White and Negro College Students in the Various Traits

Trait	White S. D.	Negro S. D.	White Excess	Per Cent	Negro Excess	Per Cent
Speed of Movement	2.531	3.070			.539	21.3
Freedom from Load	2.537	2.716			.179	7.0
Flexibility	1.740	2.000			.260	14.9
Speed of Decision	2.888	3.045			.157	5.4
Motor Impulsion	1.781	1.955			.174	9.7
Reaction to Contradiction	2.451	1.997	.454	22.2		
Resistance to Opposition	2.731	2.641	.090	3.4		
Finality of Judgment	2.392	2.301	.091	3.8		
Motor Inhibition	3.409	3.519			.110	3.2
Interest in Detail	1.489	2.146			.657	44.1
Co-ordination of Impulses	2.882	2.838	.044	1.5		
Volitional Perseveration	2.965	2.524	.441	18.2		
Average Per Centage of Excess				9.8		15.1

After the comparisons given above were made, the results were analyzed in terms of ages and academic grades. Except for the preponderance of freshmen in the white high school group, and

---

[27] Mayo, M. J. *The Mental Capacity of the American Negro.* Archives of Psychology, Monograph No. 28, 1913.

[28] Derrick, S. M. *A Study of the Intelligence of White and Colored College Students.* Journal of Applied Psychology, Vol. 4, pp. 316-329, December, 1920.

TABLE III.

Distribution of subjects by ages and grades.

WHITE STUDENTS

	High School				College			
Age	Fresh.	Soph.	Jun.	Sen.	Fresh.	Soph.	Jun.	Sen.
10	1							
11								
12								
13	5							
14	2	1	1					
15	5	1	1					
16	3			1				
17	1	1		2	1			
18	2	1	1	1	2	5		
19	1				1	6	3	1
20	1	1				5	3	
21		2	1			3	4	
22		1				2		
23	1					1		
24							1	
25								
26						1		
Total	22	8	4	4	4	23	11	1

NEGRO STUDENTS

	High School				College			
Age	Fresh.	Soph.	Jun.	Sen.	Fresh.	Soph.	Jun.	Sen.
12	1							
13								
14	2							
15		1						
16	1							
17	2				1			
18	3	5	1	1	1			
19	2	2	1	1	1	2		
20		1	2	1	3	2		
21			1	1	2		2	1
22						2	1	1
23	1		1	3	1	1		2
24			2		1	1	2	2
25			1			1	1	
26				1			1	2
27					1			3
28								1
Total	12	9	9	8	11	9	7	12

TABLE IV.

Mean scores by ages of subjects.

WHITE STUDENTS

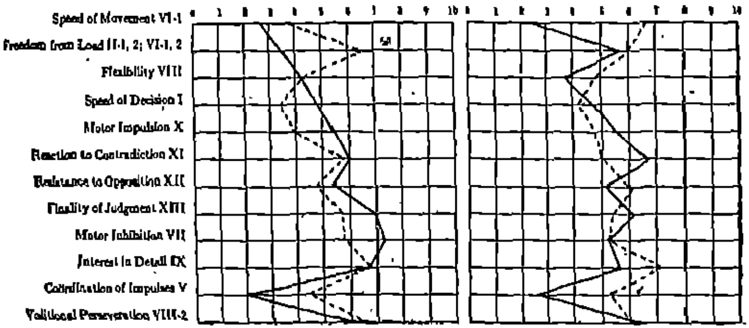
Age	SM	FL	F	SD	Imp.	RC	RO	FJ	Inh.	ID	CI	VP
10-16	3.81	6.38	4.24	3.38	3.90	5.81	4.81	5.66	5.76	6.76	4.38	6.28
17-19	6.65	6.07	4.89	4.20	4.79	5.00	6.14	5.38	5.24	7.14	5.20	6.07
20-22	5.43	5.82	4.78	3.48	3.69	5.91	5.69	4.78	4.17	6.82	4.56	6.52
23-28	5.75	5.75	5.50	2.25	3.25	7.50	6.00	5.00	3.75	6.25	5.75	5.50

NEGRO STUDENTS

Age	SM	FL	F	SD	Imp.	RC	RO	FJ	Inh.	ID	CI	VP
10-16	2.00	3.40	4.20	4.80	5.40	6.00	5.40	7.00	7.20	6.80	2.00	6.00
17-19	2.35	5.74	3.87	4.66	5.49	6.69	5.17	6.09	5.26	5.65	2.69	6.22
20-22	5.15	6.30	4.10	5.00	4.85	6.10	5.00	5.95	4.50	6.95	4.00	5.40
23-28	3.62	5.68	4.31	3.24	5.55	6.48	5.31	4.55	5.52	5.34	2.76	5.79

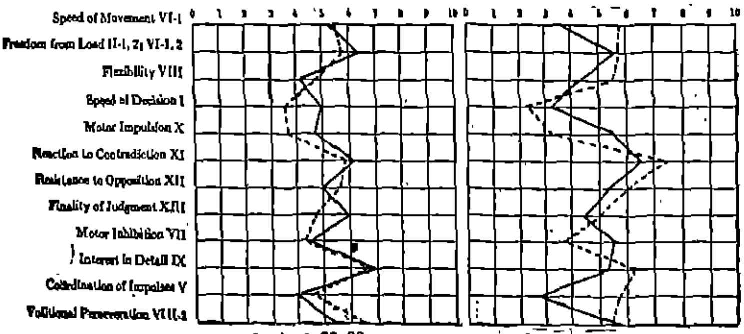
For number of subjects in each group, and grades represented, see Table 3.

Figure IV.



a. Ages 10-16

b. Ages 17-19



c. Ages 20-22

d. Ages 23-28

of sophomores in the white college group, the distribution of students is approximately the same in both races. Table III shows the distribution by age and grades. From inspection of this table, it is seen at once that the negro student is from two to three years older than the white student of the same grade. This retardation may be caused by adverse environment or by inferior intelligence of the negro. Probably both factors are at work. To eliminate as much as possible the influence of age on the scores made by the groups, the subjects were divided into four sections, the first section containing those 16 years of age and younger, the second section containing those from 17 through 19, the third those from 20 through 22, and the fourth those 23 and above. Table IV gives the mean scores made by each section in each trait and Figure 4 shows the profiles plotted from these data. It will be seen from inspection that, as the subjects grow older, there is a tendency for increased scores in speed of movement, flexibility, reaction to contradiction, and co-ordination of impulses; a tendency for decreased scores in speed of decision, finality of judgment, motor inhibition, and perhaps volitional perseveration; the scores for motor impulsion tend to remain about the same; in freedom from load, the white scores remain about the same, or decrease slightly, while the negro scores increase with age; in resistance to opposition, the white scores increase, and the negro scores remain about the same; in interest in detail, the white scores decrease slightly, while the negro scores vary. The increase in speed of movement may be due to increased use of writing, the necessity for rapid note-taking, or the influence of other situations in every-day school life that call for speed in writing. The increase in the three other traits noted would probably be due to a deeper realization of problems as such, and the consequent tendency to eliminate "snap judgments". The decrease in motor inhibition and volitional perseveration may be due to the greater value which the older person puts upon time, or, in the case of motor inhibition, the decrease of content which the test-phrase holds for him, and the consequent increase in difficulty of the test. A person's rating in motor impulsion is probably a fundamental part of his make-up, and would hardly be expected to change in either direction. A true race difference may be indicated by the difference in the change of scores for freedom from load, in which those for the whites remained nearly constant, and those for the negroes' increased. The negro score in this trait never increased enough, however, to bring it up to the lowest score of the whites. In such a trait as resistance to opposition, the increase in white scores, while the negro scores remain nearly constant, may indicate an earlier maturity on the part of the negro. The decrease in scores for the white students in interest in detail may indicate, as noted above, an increased valuation of time.

TABLE V.

Mean scores by school grades of subjects.

WHITE STUDENTS

Grade		SM	FL	F	SD	Imp.	RC	RO	FJ	Inh.	ID	CI	VP
HS	1-2	3.56	6.46	3.93	2.76	3.70	5.60	5.50	5.36	4.36	6.16	3.90	5.66
HS	3-4	6.25	7.62	4.87	4.87	4.00	5.12	5.60	6.50	5.60	7.50	3.62	5.85
C	1-2	7.04	5.74	5.04	4.00	4.33	6.89	5.74	4.88	6.96	7.22	6.44	6.58
C	3-4	6.08	7.66	5.41	4.33	4.91	4.83	5.83	5.00	4.17	7.58	5.08	6.75

NEGRO STUDENTS

Grade		SM	FL	F	SD	Imp.	RC	RO	FJ	Inh.	ID	CI	VP
HS	1-2	2.09	5.09	3.38	4.33	5.57	6.47	5.00	6.81	6.38	4.91	2.33	5.71
HS	3-4	4.06	6.82	4.23	5.18	5.35	7.23	5.12	5.65	7.23	7.29	1.88	5.88
C	1-2	4.40	5.25	4.05	4.75	4.95	5.85	5.65	5.20	5.45	6.15	3.85	5.30
C	3-4	3.85	5.79	3.95	2.95	4.84	6.37	5.42	4.21	3.79	7.58	3.89	6.10

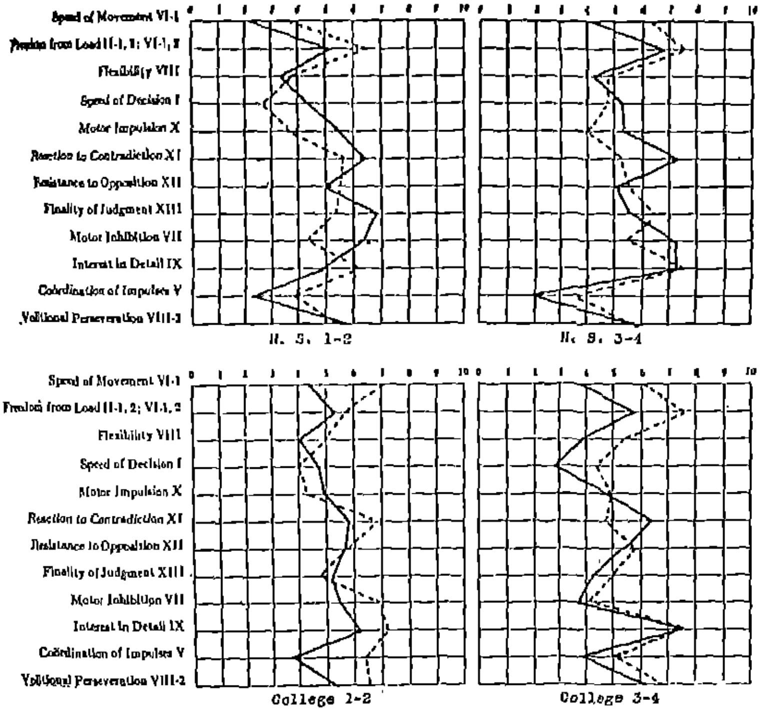
Average ages of above groups:

	HS 1-2	HS 3-4	C 1-2	C 3-4
Whites	16.3	17.0	19.74	20.33
Negroes	17.66	21.59	21.2	25.58

For distribution of ages, and numbers in each group, see Table 3.

In order to check the above classification of the groups into age-sections, another classification was made, this time into grade-sections. Here it is evident, the factors of intelligence and greater capability come into play, and we may be fairly sure that the higher the grade, the higher the intelligence level, since there is a selective factor at work to eliminate those of less native capacity. Table V gives the mean scores attained by the various sections, and Figure 5 shows the profiles plotted from the data. In this case, with increase in grade, there is an increase in the scores of more traits than when the student's increase is in age. Inspection shows that there are increases in the scores for speed of movement, flexibility, resistance to opposition, interest in detail, co-ordination of impulses and volitional perseveration. Finality of judgment is the only trait in which there is a decrease. The scores for freedom from load, reaction to contradiction, and motor inhibition remain about the same, or vary irregularly. There is a slight increase in the scores for speed of decision and motor impulsion for the whites, with a decrease in these traits for the negroes. With increased intelligence and stronger personality, one would naturally expect higher scores. For finality of judgment, the decrease is from above "5" to about "5" which is the normal score. The advanced students seem to be less addicted to hasty revision. The constancy of the traits freedom from load, reaction to contradiction, and motor inhibition, may indicate that normal scores in these traits are sufficient for proficiency in school work,

Figure V



although it would seem that the advanced student would gain in resistance to contradiction. The opposite direction in which the scores tend, for the two races, in speed of decision and motor impulsion is probably indicative of racial characteristics, and may offer good returns for further investigation.

Since, as stated in the first part of this section, a study of mean or median scores offers little that is definite in the way of racial temperament patterns, the scores for the college students were gone over in an attempt to list any tendency toward groupings of the high scores found there. Following Professor Downey in the main, we may suggest that there are three types of such groupings—those in the traits listed as the first four characterizing one as mobile or rapid-fire in organization, as the second four characterizing him as aggressive or forceful, and as the third four the controlled, deliberate, careful type. Also, there may be combinations of any two groupings, or of none at all—the profile may be balanced or else very irregular. The method followed here in listing tendencies was as follows: The first four scores in each individual's case were added together, as were the second and third sets of fours. These sums were compared, and when the sum of one group exceeded the sum of either of the other groups, the individual was listed as being characterized by the tendency correlated with that group. Where no one group exceeded either of the others by three or more points, there were two possibilities—either that the individual had a well balanced profile, or else there was a combination of two high groups and one low one. If the first was the case, he was listed as "Balanced"; if the second was the case, he was listed under the combination of his two high groups. Table VI gives the results of this attempt to classify the patterns. From this table we may see that the whites slightly surpass the negroes in the number of mobile, rapid-fire individuals; they have a clear superiority in the number of controlled, deliberate, careful persons; and they slightly surpass the negroes in the number of individuals showing a combination of these two characteristics. The negroes slightly surpass the whites in the number of aggressive persons and in the number of persons combining quickness and mobility with aggressiveness, and also in the number of individuals combining aggressiveness and deliberation.

TABLE VI

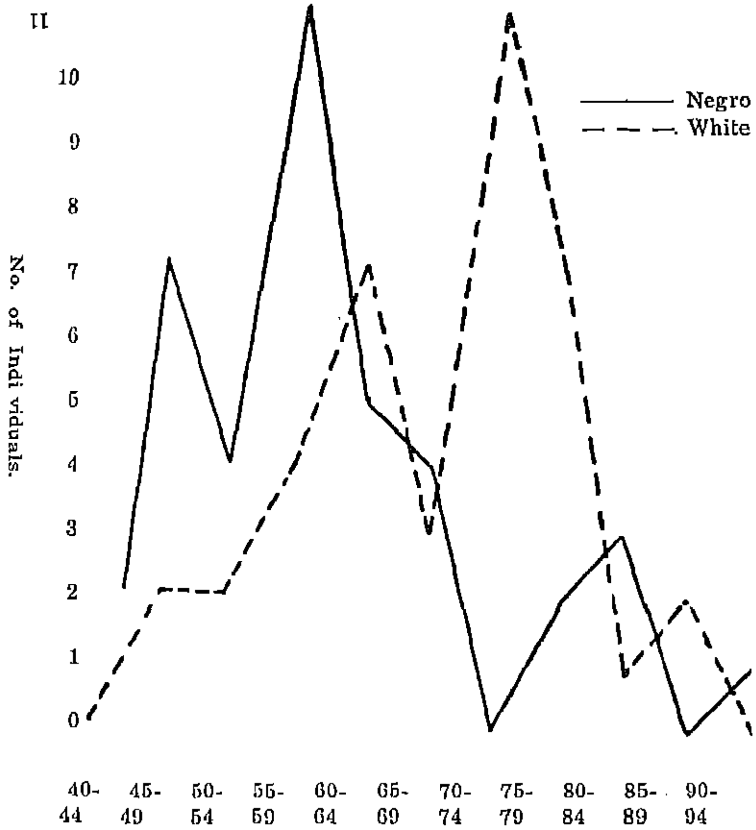
Showing distribution of college students according to the group of traits in which each received his highest scores.

	1st	2nd	3rd	1st and 2nd	2nd and 3rd	1st and 3rd	Balanced
Whites ...	4	6	14	1	6	5	3
Negroes	2	9	8	4	8	3	5

As a last point, Professor Downey suggests that, while there is not as much significance attached to the total score of each individual as to his scores on various traits, still, a high total score is, in general, indicative of a "stronger personality," and a low total score is indicative of a weaker personality. On this assumption, a tabulation of total scores of the college group was made, and a distribution curve was drawn from the results. See Figure 6.

FIGURE 6.

Distribution of total scores of College students.



	Mean	Median
White	68.08	70
Negro	59.05	57

Six negroes (15.4%) passed the white median.

It was found that the white college students surpassed the negro college students in their totals, which, therefore, would indicate, that, in general, the whites have "greater force of personality."

SUMMARY

1. In general, the negro is slower in movement than the white, has about the same or slightly greater load or inertia; has slightly less flexibility; is slightly quicker in making his decisions; his motor impulsion is slightly greater; reacts more firmly against contradiction; but offers less resistance to physical opposition; takes about the same or slightly less time in reconsidering his decisions; the negro college student has less motor inhibition, the negro high school student has greater, compared with white students; the negro has practically the same interest in detail as the white, has markedly less ability in co-ordinating his impulses; and has a slightly less tendency to volitional perseveration.

2. There is little difference in variability, but the negro college student has greater variability on seven of the traits, as opposed to the greater variability of the whites in five, and the amount of the negroes' variability is slightly greater.

3. In general, the negro college and high school student is from two to three years older than the white student of the same academic grade.

4. As to the influence of age (from early adolescence to the middle twenties, as the subject approaches maturity, there is a tendency for increased scores in speed of movement, flexibility, reaction to contradiction, and co-ordination of impulses; a tendency for decreased scores in speed of decision, finality of judgment motor inhibition, and volitional perseveration, the scores for motor impulsion tend to remain about the same; in freedom from load, the white scores remain about the same, or decrease slightly, and the negro scores increase with age; in resistance to opposition, the white scores increase and the negro scores remain about the same; in interest in detail, the white scores decrease, while the negro scores vary.

5. As to the analysis of mean scores by grades (probably influenced by the factor of increasing intelligence in direct ratio to the higher grades): the scores for speed of movement, flexibility, resistance to opposition, interest in detail, co-ordination of impulses, and volitional perseveration increase in the higher grades; the scores for freedom from load, reaction to contradiction, and motor inhibition vary; there is a slight increase in the scores for speed of decision and motor impulsion for the whites, with a decrease in these traits for the negroes, from the lower to the higher grades.

6. In predominance of temperamental patterns, the whites slightly surpass the negroes in the number of mobile, rapid-fire individuals; they have a clear superiority in the number of con-

tròlled, deliberate, careful persons; and they slightly surpass the negroes in the number of individuals showing a combination of these two characteristics. The negroes slightly surpass the whites in the number of aggressive persons and in the number of individuals combining quickness and mobility with aggressiveness, and also in the number of individuals combining aggressiveness and deliberation.

7. In "strength of personality," as indicated by the total score of each individual, the whites surpass the negroes, only 15.4 per cent of the negroes exceeding the median of the whites.

A STUDY OF THE DATA ON THE RESULTS GATHERED FROM REPEATED MENTAL EXAMINATIONS OF 200 DEFECTIVE CHILDREN ATTENDING SPECIAL SCHOOLS OVER A PERIOD OF EIGHT YEARS

By Meta Anderson

Director Binet Classes, Newark, N. J.

Although the work for defectives was begun in Newark in September, 1910, it was not until 1913 that the organization of the work permitted anything like systematic re-examination of the children admitted to the special schools and even then complete yearly re-examination of the whole special school population was not accomplished. However, since 1916 it has been possible to re-examine each child in the special schools yearly excepting a few cases which will be noted later. The object of the work was to get information in regard to the growth of intelligence of the children in the special classes (now called Binet Classes) in order, if possible, to know better how to train them and to eliminate errors in their training and education. The object of this study was to organize, classify and analyze these data in order to get further information in regard to types of mental growth curves and in order to get further information in regard to the value of the intelligence quotient.

While the population of the Binet Schools is upward of five hundred pupils only two hundred were studied here. These two hundred cases vary from the grade of low grade imbecile to that of high grade moron. Before January, 1918, the cases were selected for the Binet schools because they were failures in the regular grades and at the same time fell more than three years below normal according to an intelligence test (this standard being set by the New Jersey law). Since January, 1918, the cases have been selected by the Psycho-Educational Clinic of the city. Each case is certified as being definitely feeble-minded before being sent to a Binet School.

Each case was examined from three to eight times at not less than a year's interval. The majority of the cases were re-examined by the writer. The others were examined by two head assistants in the department both of whom were unusually well trained to do the work.

In a department always having a population of 500 defective children it would seem possible to have numbers of retests on at least that many children. This investigation has shown that the turnover in the population of the Binet Schools is very great.

Many of the children except the lowest grade are assigned to special work late in their school career. These children are not presented for psychological examination until the regular schools have done all in their power to help them succeed in their work. Hence, it is only after their inability to learn has become most marked or their behavior disorders have become unbearable that these children are assigned to the Binet Schools. By that time the children are anywhere from ten to fourteen years of age. Since they leave school at sixteen and sometimes earlier as during the war when the pressure for workers was very great these children have short opportunity for training. This also renders it impossible to have a great number of retests on the same individual.

A certain percentage of this group also leave school to go to houses of correction, to institutions, move from the city or are lost track of. Therefore the number remaining in the Binet Schools for more than four years is comparatively few.

With the present wider use of group tests defective children are being discovered earlier in their school career but it will be necessarily a number of years before sufficient data are gathered on this new group for a worthwhile study.

Before 1918 the Goddard revision of the Binet tests were used, since 1918 the Stanford revision of the Binet tests have been used.

The study falls into two parts. The first part has to do with the frequency of types of mental growth curves and the second part has to do with a study in values of the IQ.

PART I. FREQUENCY OF TYPES OF MENTAL GROWTH CURVES

OTHER INVESTIGATIONS.

Kuhlman has made growth curves on data gathered from repeated retests on inmates of the Minnesota institution and, he says: "The main features of these growth curves present but little that is not now pretty well known. They show that the feeble-minded do develop mentally, and at a rate in proportion to the grade of intelligence. This was still a disputed point at the time this study was begun. This rate of development decreases with age for all grades. The rate of this decrease cannot, of course, be shown with this data, as we do not know how much mental growth in terms of absolute units the mental year represents at different age levels."

Kuhlman summarizes his results on mental age growth as follows:

"1. The mental ages of the feeble-minded increase with age at a rate proportionate to the degree of mental deficiency.

"2. On the whole the mental age ceases to increase between the ages of fifteen to eighteen, the idiot grade ceasing development about three years earlier than the borderline grade.

"3. For the cases figuring in the present study and for the time period they were under observation, 4.8 per cent of the cases gained twelve months or more in mental age in a year, 68 per cent gain from one to eleven months a year, 11 per cent gain or lose less than one month a year, and 16 per cent lose from one to seven months a year in mental age.

"4. The lower grades lose more frequently in mental age than the higher grades.

"5. The frequency in loss in mental age increases with age, independently of grade."[1]

Doll says on the rate of growth for subnormals: "Not much more is known about the rate of mental growth among extreme subnormals, like the feeble-minded, than about the age of arrest. In this regard we are even at a loss for adequate observational data. Experience with institutional feeble-minded, however, suggests a great variety of possible forms of retardation curves. Such experience points to a small but significant group of feeble-minded who grade 'at age' by the Binet scale when first examined but who make no material progress in later examinations over a period of years (Type A.) There are others (Type B) who at the time of first examination test at age, then continue to develop at a retarded rate for a time and then reach a point of arrest. There is a third group (Type C) already retarded when first observed, who from the first examination to the age of final arrest develop at an annual rate of growth which is a constantly decreasing variable. There is also a fourth group (Type D) who from the time of first examination to the age of final arrest. There are many variants and combinations of these basic types for individual subjects. We believe, however, that these four types comprise the majority of the feeble-minded. It is conceivable, though not probable, that some individuals might have a higher than average rate before retarding. It is also conceivable, though not likely, that the arrest might be relatively greater in the early ages with a curve of growth which would be accelerating at some intervals. It remains for experimental investigation to show the relative frequencies of these different types of growth curves."[2]

---

[1] Kuhlmann, Results of Repeated Re-examinations of Feeble-minded (639) Over a Period og Ten Years. Journal of Applied Psy. Sept., 1021.

[2] Doll, Growth of Intelligence. Psy. Monograph, pp. 15-10.

With the view of determining what light, if any, the limited data on hand could throw on the subject of the growth of intelligence of feeble-minded subjects 100 mental age growth curves of pupils attending the Binet Schools from three to eight years were plotted. These curves were then classified according to Doll's theoretical intelligence growth curves as follows:

Type A. (Doll) Those who test "at age" by the Binet scale when first examined but make no material progress in later examinations.

Type B. (Doll) Those who at the time of first examination test "at age," then continue to develop at a retarded rate for a time and then reach a point of arrest.

Type C. (Doll) Those who are already retarded at time of first examination but who continue to develop from the time of first examination until time of final arrest at an annual rate of growth which is a constantly decreasing variable.

Type D. (Doll) Those who from the time of first examination to the age of final arrest develop at a constant rate up to time of final arrest.

For convenience two new types were added although it is quite likely they could be easily included in one of the above groups. They are as follows:

Type E. (New) Those who are retarded at the time of first examination, and who also have apparently reached the time of final arrest.

Type F. (New) Those who have developed at a constant or variable rate up to the time of apparent arrest, but who after two to four years of apparent arrest, have begun to develop again.

NUMBER OF CASES	TYPE
1	4 to 5
1	A
0	B
47	C
22	D
19	E
11	F
100	

Of the 11 cases listed as *type f* there were

4 who remained at an apparent arrest for 2 years.
5 who remained at an apparent arrest for 3 years.
2 who remained at an apparent arrest for 4 years.

Of these 11 cases listed as *type f* there were

2 cases who continued development for 1 year after arrest.
6 cases who continued development for 2 years after arrest.
2 cases who continued development for 3 years after arrest.
1 case who continued development for 4 years after arrest.
at which time there was another apparent arrest or the children left school.

Type f cases classified according to life age

Number of Cases	Life Age
1	6 to 7
1	7 to 8
5	9 to 10
1	10 to 11
2	11 to 12
1	12 to 13
11	

Type f cases classified according to mental age

Number of Cases	Mental Age
2	3 to 4
6	6 to 7
1	7 to 8
1	8 to 9
11	

SUMMARY.

1. It is evident that a greater number (47) of the cases studied in this group have developed mentally according to Doll's theoretical curve described as Type C.

2. The next larger group (22) of cases studied have developed mentally according to Doll's theoretical curve described as Type D.

3. There were 67 cases out of the 100 studied who were retarded when first examined and who then retarded either at a constant rate or at a steadily decreasing rate up to age of final arrest.

4. Only one (1) case could be obviously classified under Type A.

5. The 19 cases which were retarded when first examined and which developed no further at later examinations may or may not belong to Type A. That is, it is possible that if they had been examined at earlier periods they would have tested "at age."

6. The 11 cases under Type F could be classified under Type C in a broad classification. They differ from that type merely in the fact of having long periods of arrest but in the long run they develop at a decreasing rate up to the time of final arrest.

7. It would appear from the study of these 100 cases that the majority 80 in this study of feeble-minded cases slow up gradually in mentality up to time of final arrest and that comparatively few develop normally for a time and then stop suddenly. *In this study not more than 20 of the cases could have* been said to have come to sudden arrest and 19 out of the 20 cases are doubtful.

8. The 100 cases studied were all classified as feeble-minded and were *attending* Binet Schools. *It* is possible that feeble-minded cases which develop normally and then retard suddenly are not discovered early enough in their school careers to be selected for Binet Schools.

PART II. STUDY IN SOME VALUES OF THE I. Q.

From three to seven tests were made on 200 children attending the Binet Schools at intervals of a year in most cases. There are cases which are noted in which the retests were made at greater intervals because the children were not available for examination either because they had moved from the city or attended another school and then re-entered the Binet School when they were re-examined. At the time of this study there were *only* two hundred *cases* who had been re-examined a sufficient number of times to make the study of their cases worth while. These cases then were all selected because they were attending the Binet Schools and because they had attended long enough to have from two to six retests made at intervals of a year.

The cases were classified and grouped.

(a) According to life age at initial test and I. Q. in groups of tens.

(b) According to I. Q. in groups of tens.

(c) According to I. Q. and intervals of retests regardless of life age.

These groupings were made in order to study the I. Q. in its various aspects.

A. CASES CLASSIFIED ACCORDING TO LIFE AGE AT INITIAL TEST AND I Q—IN GROUPS OF TENS.

No. of Cases.	Life Age.	Median I. Q. at First Test.	Number of Retests.	Median I. Q. at Last Test.
0	6			
2	7	38½	7	44½
1	8	39	6	41
1	9	33	4	44
2	10	31	4	25
0	11			
0	12			
1	13	36	4	29
7				
1	6	44	6	42
0	7			
3	8	42	6	44
3	9	47	4	47
0	10			
2	11	44	3	48
1	12	49	5	40
0	13			
10				
3	6	55	4	44
1	7	58	5	58
3	8	59	5	55
7	9	54	3	49½
4	10	53½	3	51½
6	11	54½	4	51
2	12	57	3	48½
26				
0	6			
1	7	69	3	57
3	8	61	4	53
15	9	67	5	57
5	10	67	4	58
8	11	64½	3	57
6	12	64	3	57½
2	13	64	3	56
40				

A. CASES CLASSIFIED ACCORDING TO LIFE AGE AT INITIAL TEST AND I Q—IN GROUPS OF TENS. (CONTINUED)

No. of Cases.	Life Age.	Median I. Q. at First Test.	Number of Retests.	Median I. Q. at Last Test.
2	6	74	4	60½
1	7	70	7	60
2	8	75	4	65
6	9	74	6	48
5	10	73½	5	56½
8	11	73½	3	60
5	12	72	3	60
29				

SUMMARY OF RESULTS.

1. *The cases with I. Q.'s from 30 to 39, inclusive, and with life* ages from six to nine years, inclusive, showed a gain in median I. Q. of 2 to 11 points.

2. The cases with I. Q.'s from 30 to 39, inclusive, and with life ages from 10 to 13, inclusive, showed a loss of 6 to 7 points in median I. Q.

3. *The cases with I. Q.'s from 40 to 49, inclusive, and with life* ages of 6 years showed a loss of 2 points in median I. Q.

4. The cases with I. Q.'s from 40 to 49, inclusive, and with life ages from 8 to 11, inclusive, showed a gain in median I. Q. of 2 to 4 points.

5. The cases with I. Q.'s from 40 to 49, inclusive, and with the life ages of 12 showed a loss in median I. Q. of 5 points.

6. The cases with I. Q.'s from 50 to 59, inclusive, and with life ages from 6 to 12, inclusive, showed a loss in median I. Q. of nothing to 9 points.

7. The cases with I. Q.'s from 60 to 69, inclusive, and with life ages from 6 to 13, inclusive, showed a loss in median I. Q. of 7 to

2. The cases with I. Q.'s between 40 and 49 showed a loss of 1

12 points.

8. The cases with I. Q.'s from 70 to 79, inclusive, and with life ages from 6 to 12, inclusive, showed a loss in median I. Q. of 10 to 26 points.

9. In other words, the younger cases with I. Q.'s from 30 to 39. inclusive, gained in median I. Q.; the older cases with I. Q.'s from *30 to 39, inclusive, lost in median I. Q.*

10. With the exception of the 6-year-olds who lost 2 points in median I. Q. the younger cases with I. Q.'s from 40 to 49, inclusive,

gained in median I. Q.; the older cases with I. Q.'s from 40 to 49, inclusive lost in median I. Q.'s.

11. All cases of any age with I. Q.'s from 50 to 80, inclusive showed a loss in median I. Q.

D. CASES CLASSIFIED ACCORDING TO IQ

No. of	I. Q. Range	Median I.Q.'s at Successive Examinations.					
Cases		1	2	3	4	5	6
14	30-39	35	33	31	30	37	
29	40-49	47	49	46^{2}	48	46	
63	50-59	56	56	54	51^{2}	50	48
59	60-69	65	60	60	58	60	57
33	70-79	73	67	62	60	58	56

SUMMARY.

1. The cases with I. Q.'s between 30 and 39 showed a gain of 2 points in median I. Q. at the 5th re-examination.
point in median I. Q. at the 5th re-examination.

3. The cases with I. Q.'s between 50 and 59 showed a loss in median I. Q. of 8 points at the 6th re-examination.

4. The cases with I. Q.'s between 60 and 69 showed a loss in median I. Q. of 8 points at the 6th re-examination.

5. The cases with I. Q.'s between 70 and 79 showed a loss in median I. Q. of 17 points at the 6th re-examination.

6. All cases with I. Q.'s below 50 remained practically stationary with a gain or loss of but 1 or 2 points in median I. Q.

7. The cases with I. Q.'s between 70 and 80 showed loss. The highest initial I. Q. showed the greatest loss in median I. Q. at the final re-examination.

C. CASES CLASSIFIED ACCORDING TO IQ AND ALSO ACCORDING TO INTERVALS OF RETESTS. •

No. of cases	IQ range	Median IQ's at initial test and at intervals of retests Compared.					
		1	2	3	4	5	6
12	30-39	34-32^1					
10	30-39		33^2-28^2				
12	30-39			34-33			
9	30-39				34-37		
5	30-39					39-42	
33	40-49	46-48					
24	40-49		46-45^3				
20	40-49			.44-48			
13	40-49				44-42		
7						44-43	
67	50-59	50-56					
54	50-59		55-53				
40	50-59			55^1-56^2			
19	50-59				55-50		
12	50-59					53^2-48	
7	50-59						56-49
74	60-69	65-61					
64	60-69		65-60				
43	60-69			65-58			
21	60-69				67-60		
9	60-69					65-56	
33	70-79	73-67					
31	70-79		73-62				
25	70-79			73-61			
16	70-79				72-58		
11	70-79					72-56	

SUMMARY.

All median IQ's in all ranges showed a loss as follows:

Range of IQ.	Points lost at successive examinations.					
	1	2	3	4	5	6
30-39	1	2	3	3	5	
40-49	½	1	2	2	4	
50-59	0	1	3	4½	5	
60-69	4	5	7	7	9	
70-79	6	8	11	14	16	

In each group of cases the greater loss in median IQ occurred at the end of the fifth examination.

In each group of cases the greatest loss in median IQ occurred in the groups of the higher initial IQ's.

CONCLUSIONS.

1. A great majority of the cases who are examined as candidates for the Binet Schools are retarded at the time of first examination and then retard either constantly or at a decreasing ratio up to time of final arrest.

2. Comparatively few of the cases examined as candidates for the Binet Schools develop normally up to a certain age of arrest and then remain at arrest.

3. In cases whose Intelligence Quotient is below 50 there is a tendency for the median IQ to remain constant or to gain or lose slightly. The median IQ shows a gain among the younger cases.

4. All cases with IQ's between 50 and 80 show a decrease in median IQ at any age. The highest initial IQ's show the greatest loss in median IQ after repeated examinations.

5. In the case of examinations at varying intervals the greatest loss in median IQ was at the interval of five years which was the last examination for all grades of cases. The greatest loss in median IQ was shown by the highest grade of cases.

6. The IQ is variable in individual cases.

7. No group of cases showed a marked increase in median IQ.

8. These conclusions are merely for cases examined as candidates for Binet Classes. It is possible that repeated examinations on all retarded cases in a school system might give different results.

THE DOWNEY WILL-TEMPERAMENT GROUP TEST; A FURTHER ANALYSIS OF ITS RELIABILITY AND VALIDITY

By G. M. Ruch and M. C. Del Manzo,
University of Iowa.

This paper presents the results of a statistical analysis of the scores of one hundred and forty-six high school students tested with the *Downey Group Will-Temperament Test.* The discussion of these results will be divided into four sections in accordance with the specific objectives of the investigation.

I The Reliability of the Test as a Whole

Since the Downey tests are published in only a single form the usual method of obtaining reliability coefficients by correlating equivalent forms is not applicable here. Moreover, the common practice of breaking tests into chance halves for purposes of obtaining the self-correlation can only be carried out with certain qualifications as to the justice of this method. In an earlier paper[1], one of the authors has argued that the correlation of the sums of the odd-numbered tests against the sums of the even-numbered tests might fairly provide an approximation to a valid coefficient of reliability. A few words in explanation of the nature of the *Downey Will-Temperament Tests* will help to make clear the limitations of the method of correlating chance halves.

The scale is composed of twelve separate tests each designed to measure one aspect of the total volitional pattern. These are:

A Speed of Movement (VI-1)
B Freedom from Load (VI-1, 2)
C Flexibility (VIII)
D Speed of Decision (I)

E Motor Impulsion (X)
F Self-Confidence (XI)
G Non-compliance (XII)
H Finality of Judgment (XIII)

I Motor Inhibition (VII)
J Interest in Detail (IX)
K Co-ordination of Impulses (V)
L Volitional Perseveration (VIII-2)

---

[1] Ruch, G. M.: "A Preliminary Study of the Correlations between Estimates of Volitional Traits and the Results from the Downey 'Will-Profiles'," Journal of Applied Psychology, 1921, 5, 159-182.

This arrangement of the twelve traits into three groups as here indicated follows the purpose of the author of the test in that traits A to D belong to a common "pattern"; traits E to H belong to a second "pattern"; and traits I to L belong to a third "pattern." The groupings are intended to throw together the tests that are intended to reveal, respectively, three main types of volitional temperaments known, in order, as:

I "Hair-trigger" organization, or the explosive type of will described by James (Tests A, B, C, and D).
II "Willful, aggressive pattern" (Tests E, F, G, and H).
III "Slow, accurate, tenacious type" (Tests I, J, K, and L).

The grouping and characterizations of the author of the tests is adhered to here strictly.

Since the scale is of the nature of a diagnostic picture in which certain groups (the three patterns) of tests can logically yield uniformly high scores, while other portions can as logically yield very low scores (or at least much lower scores), it follows that the total score on the whole scale will have very little meaning. It also follows that any attempt to split the total scale into chance halves must take care that equal numbers of tests be taken from each of the three diagnostic patterns in forming such chance halves. If this is done properly we would then have two half-scales; each half-scale being composed of two tests from each of the three patterns or a total of six tests in each half scale. Such chance halves could be made up in this particular case in several way, viz., (1) the grouping of all of the odd-numbered tests in one half-scale and all of the even-numbered tests in the other half-scale, (2) the grouping of the first two tests of every pattern into one of the half-scales and the last two tests of each pattern into the other half-scale, or, (3) the grouping of the first and last tests of each pattern into one half-scale and the second and third tests of each pattern into the second half-scale.

In the paper mentioned above, the reliability of the Downey tests (individual form) was studied only by the first of these three methods. In the present report, the reliability coefficients have been determined by all of the three possible ways and the results averaged. Table 1 shows these results. The notations as "Method 1," "Method 2," and "Method 3" refer to the above-mentioned ways of obtaining chance halves. The notation "$r_{\frac{1}{2}\frac{1}{2}}$" means that the coefficient of correlation stated was obtained by correlating half (six tests) of the scale with the remaining half (six tests) of the scale. The notation "r_{11}" is the porphecied correlation which would be obtained if two similar whole forms were correlated. The values of the "r_{11}" column were calculated by the use of Brown's formula, $r = \dfrac{n\ r}{1 + (n-1)\ r}$ The ages of the subjects cover a range of from ten to twenty-nine years. Grades seven to twelve are involved.

TABLE 1

RELIABILITY COEFFICIENTS

Method	$r_{\frac{1}{2}\frac{1}{2}}$	r_{11} (Brown's formula)	N
I	0.17	0.29	146
II	0.31	0.47	146
III	0.24	0.38	146
Average	0.24	0.38	146

The probable errors of these coefficients (Pearson product-moment formula) range between 0.05 and 0.06. Table 7 shows the probable errors of coefficients of correlation for the various values of r reported in this study.

In view of the wide range of talent represented in the group tested, these reliability coefficients cannot be considered to be very satisfactory. They are, however, much higher than those published in the earlier paper cited; the range of talent in the former group being very much more curtailed.

There is another possibility which must be considered in accepting these reliability coefficients at their face value. According to our original assumptions, the reliability of the scale as a whole can only be obtained by the methods we have used *in case the tests are correctly grouped into patterns in a proper manner*. The low reliability which we have found might mean merely that the tests are improperly classified at present. In this case, the true reliability of the scale (properly arranged) might be markedly higher than we have reported. Discussion of this question of the grouping of the separate tests should be postponed until the intercorrelations of the tests have been presented.

II THE INTERNAL COHERENCE OF THE SCALE

One of the crucial issues in the evaluation of the Will-Temperament Test is the matter of the arrangement of the separate elements into the diagnostic patterns. As the authors interpret Dr. Downey's classification of the individual tests into groups which picture volitional patterns, it would seem that a valid grouping would imply certain mutual relationships among the tests. These are:

(1) High scores in one test of a pattern would be paralleled with high scores in the other three tests of the same pattern. This is equivalent to saying that the inter-correlations between tests of a given pattern would tend to be positive and high.

(2) Correlations between each test and the composites of the tests of the same pattern would tend to be positive and high.

(3) The correlations between tests of one pattern and those of the other patterns would tend to be zero or very low.

(4) Correlations between single tests and the total score on the scale would be low in all probability and not significant for purposes of evaluating the tests.

In order to test these assumptions about the proper grouping of the tests of the present scale, the correlations of each test with the composite (sum) of the other three tests of the same pattern have been computed. For obvious reasons the test used as one of the variables in the correlation must be omitted from the composite. Table 2 shows the obtained correlations.

TABLE 2

CORRELATIONS OF THE TESTS WITH THE COMPOSITES

Test	Composite	r
A	B+C+D	0.54
B	A+C+D	0.28
C	A+B+D	0.10
D	A+B+C	0.39
E	F+G+H	-0.08
F	E+G+H	-0.23
G	I+F+H	-0.13
H	E+F+G	-0.09
I	J+K+L	0.07
J	I+K+L	0.33
K	I+J+L	0.07
L	I+J+K	0.05

The average inter-correlations of the tests within a pattern is probably even a better method of testing the general validity of the grouping into patterns. Table 3 shows the results yielded by this method.

TABLE 3

INTER-CORRELATIONS OF THE TESTS OF THE SAME PATTERNS

Variable 1	*Variable 2*	*r*
A	B	0.36
A	C	0.28
A	D	0.37
B	C	–0.05
B	D	0.33
C	D	0.05
Average		0.22
E	F	–0.05
E	G	0.00
E	H	–0.02
F	G	–0.16
F	H	–0.15
G	H	0.01
Average		–0.06
I	J	0.08
I	K	0.12
I	L	0.00
J	K	0.24
J	L	0.18
K	L	–0.18
Average		0.07

General inspection of Tables 2 and 3 shows that the correlations between the separate tests are uniformly low, the second pattern (Tests E to H) being the most unsatisfactory. It would be interesting to pick out from each pattern the best tests if this could be done easily. This would properly be done by multiple correlation methods and the labor involved would not seem to be justified by the general promise of the coefficients reported. In pattern one, it would appear that Tests A and D would probably be found to be considerably superior to the other two. In pattern two, there is no promise that multiple correlation methods would reveal anything of significance. In pattern three, Tests J and K would appear the most promising.

III CORRELATION OF THE WILL TRAITS WITH GENERAL INTELLIGENCE

Table 4 shows the correlations obtained between the scores in the twelve separate will tests and the averages of two standard group tests. The latter are the Terman Group Test of Mental Ability and the Morgan Group Mental Test. The reliabilities of each of these tests is known to be well above 0.90 for ranges of talent comparable with those represented in the group considered here. The significance of the coefficients of correlations of Table 4, like those of all of the preceding tables, is obscured by the fact that the reliabilities of the separate tests are unknown and are probably very low in many cases.

TABLE 4

CORRELATIONS OF THE TWELVE WILL TRAITS WITH MEASURES OF GENERAL INTELLIGENCE

Test	r
A	0.40
B	0.19
C	0.20
D	0.16
E	–0.03
F	0.20
G	0.11
H	0.14
I	0.16
J	0.33
K	0.34
L	–0.05

None of the above correlations is high enough to lend any support to the possible criticism of the Downey tests as being measures of general intelligence. The factor of intelligence undoubtedly does enter into some of the tests but not to an extent that would invalidate the scale on this score.

The inter-correlations between the tests have also been determined independently of the influence of intelligence by means of the usual methods of partial correlation. Table 5 shows such partial coefficients. The notations "rAB.i", "rAC.i", etc., read: The correlation between traits A and B which is independent of the influence of intelligence, and similarly, for the correlation between traits A and C, and the others.

TABLE 5

THE CORRELATIONS BETWEEN THE SEPARATE TRAITS WHICH EXISTS INDEPENDENTLY OF THE FACTOR OF INTELLIGENCE (OBTAINED BY PARTIAL CORRELATIONS)

$r_{AB.i}$	0.33	$r_{EF.i}$	–0.04	$r_{IJ.i}$	0.03
$r_{AC.i}$	0.22	$r_{EG.i}$	0.00	$r_{IK.i}$	0.07
$r_{AD.i}$	0.34	$r_{EH.i}$	–0.02	$r_{IL.i}$	0.01
$r_{BC.i}$	-0.09	$r_{FG.i}$	–0.17	$r_{JK.i}$	0.14
$r_{BD.i}$	0.31	$r_{FH.i}$	–0.16	$r_{JL.i}$	0.21
$r_{CD.i}$	0.02	$r_{GH.i}$	0.00	$r_{KL.i}$	–0.18

By referring to Table 4 again it will be seen that there has been little change resulting from the use of partial correlations in eliminating the influence of the factor of intelligence. It may be concluded that whatever inter-correlation does exists between the various separate tests of the same patterns is not produced by the common factor of intelligence. The formula used in partial correlation is the usual one, following Yule's notations:

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{1-r_{13}^2}\,\sqrt{1-r_{23}^2}}$$

IV THE VALIDITY OF THE DOWNEY SCALE

The validation of any test of psychology capacities is at best a most difficult task. To this rule the Downey scale is probably no exception. For an account of the derivation of these tests the reader *is referred to the manual for use with the Will-Temperament Scale*[2].

In the earlier paper to which reference has several times been made, it was attempted to secure ratings from competent judges acquainted with the subjects on each of the twelve traits of the Downey scale. A similar procedure has been adopted in the present study with certain modifications. It was thought that the teachers of the high school pupils under study might be able to furnish a set of ratings on the twelve traits which would have value as a criterion against which to evaluate the tests proper. In adopting this plan, the authors were by no means ignorant of the many pitfalls in subjective judgments. As shown in certain of the published studies, it is possible, at times, with sufficient care, to obtain reliable subjective estimates in the sense of high self-correlation among the judges. However, high reliability, although admittedly highly desirable for such purposes, does not guarantee that a valid criterion has been built up. The question of the validity of such criteria is by far the more serious issue. The fact that ten persons rating a group of individuals for "initiative," "leadership," or "tact" agree with a reliability equal to 0.75 or higher means merely that they are rating the same thing or some other thing highly correlated with this thing, but it does not follow of necessity that the ratings correspond to the psychological functions which by hypothesis underlie the thing being rated. Teachers can furnish ratings on their pupils upon such traits as "persistence," "conscientiousness," "originality," "honesty," and a host of other traits with a fair amount of reliability, yet careful statistical analysis will reveal that little else is involved in these various separate trait ratings than the teachers' general estimates of the quality of school work being done by these pupils. In other words, almost as good results could be obtained by asking for estimates of school work and then using these, in turn, as estimates of "persistence," "originality," "conscientiousness," etc. The estimates, then, are clearly not valid except as these traits are genuinely correlated with the quality of school work and even then it is a very indirect type of measurement.

---

[2] Downey, June E.: "The Will-Profile." University of Wyoming Bulletin, Vol XVI, No. 4b, November, 1919.

On the other hand what conclusions can be drawn from the situation of judges not being able to agree upon the presence or absence of quantitative statement of psychological traits in a group of individuals known to the judges? Does failure to show self-correlation in chance halves of these sets of estimates carry with it any important implication? Does such a negative result with respect to a psychological trait prove that the trait in question has no true psychological basis? No universal answer can be given to these queries. In the case of many reputed psychological rubrics it probably does mean that there is no real evidence for assuming the reality of such traits as true mental functions. In other cases if not in most the answer is left in an undeterminate status, i.e., the experimental results are entirely non-committal in the sense that they furnish no evidence; positive or negative, upon the issue of the existence or non-existence of the trait in question. Even at this it might for all practical purposes be argued that unless competent persons are in fair agreement about the presence or absence of reputed psychological characteristics there is no *direct way of identifying such traits for purposes of study.*

With these limitations of the method of estimates in mind we can return to the discussion of the immediate problem. In the attempt to build up a criterion for purposes of correlation of the trait scores with a second measure of the same traits the high school teachers of the pupils in our present study were asked simply to nominate the ten pupils in their classes who possessed the "strongest will-power" and the ten who had the "weakest will-power." No attempt at further definition of these directions was made. The reactions expected were those of the man on the street. The only outcome which was expected was that the pooled nominations might reveal two groups widely spaced upon a scale of "will-power" in the common meaning of that term. When the nomination blanks were received it was found that certain pupils were invariably reported in the same category. A larger number were about as often placed in one group as the other and a still larger number were never once nominated for either classification. There was, then, evidence of marked agreement in many cases, the total number of complete agreements being roughly 20 per cent of the total of one hundred forty-six pupils in the high school under study. Twenty-eight pupils involving equal numbers of the sexes and equal numbers in each of the two categories (i.e., "strong-willed" and "weak-willed") were selected for further study. The agreement was perfect or practically perfect in these cases in the sense that either three or four of the four teachers acquainted with every pupil placed him in the same category and no teacher gave a dissenting vote in any of these cases. The reliability coefficient for this final group was not determined but its general nature is assuredly high by virtue of the method of forming the group. *Instructions were then sent to every teacher who had nominated* any of the pupils in the final group to rate each pupil upon each of the twelve traits of the Downey scale. The trait descriptions were taken verbatim from the author's manual for use with these tests.

The final group therefore comprised (to the best knowledge of all of the teachers of the school) the fourteen pupils of the school showing the most marked volitional characteristics and an equal number most lacking in these same characteristics. As well as could be done by estimation we had segregated the highest and lowest ten per cents of the group as a whole. Table 6 shows the correlations found between the pooled estimates of the teachers and the test scores in each of the twelve traits. Inspection of these coefficients of correlation does not reveal very definite evidence of the ability of teachers to identify the various traits tested by the Downey scale. Whether this situation can be interpreted as meaning that the Downey scale is of doubtful validity is very questionable. A saner conclusion would be that teachers are unable to identify the traits of the Downey scale in their pupils to any considerable extent (traits J and K, and possibly traits A and C being apparent exceptions). It is probably fair to assume, however, that the Downey tests cannot be validated by the method of subjective estimates with any high amount of assurance. The correlations between estimates and test scores are in general of about the same magnitudes as those reported in the earlier study (where the range was from –0.33 to 0.51 for the pooled estimates of three or four college professors; the average being 0.07). The average estimate-test correlation is 0.15 for Table 6.

TABLE 6

CORRELATIONS BETWEEN ESTIMATES AND TEST SCORES (N=28)

Trait	r	Trait	r	Trait	r
A	0.37	E	0.19	I	–0.09
B	0.02	F	–0.03	J	0.51
C	0.30	G	–0.04	K	0.53
D	–0.29	H	0.17	L	0.17

The probable errors of the coefficients of correlation in Table 6 can be found in Table 7 to a sufficient accuracy for our purposes. The same is true of the correlation coefficients given in Tables 1 to 5. Except in Table 6, the numbers of cases have invariably been one hundred and forty-six.

TABLE 7

A REFERENCE TABLE OF PROBABLE ERRORS OF PEARSON COEFFICIENTS OF CORRELATION FOR VALUES OF r APPEARING IN THE FOREGOING TABLES

r	Probable Erors of r	
	N=146	N=28
0.00	.056	.127
0.05	.056	.127
0.10	.055	.126
0.15	.055	.125
0.20	.054	.122
0.25	.052	.120
0.30	.051	.116
0.35	.049	.112
0.40	.047	.107
0.45	.045	.102
0.50	.042	.096
0.55	.038	.089

CONCLUSIONS.

The following points are suggested as summarizing the more important findings presented:

(1) The reliability of the *Downey Group Will-Temperament Test* is probably not higher than 0.40 as an estimate of the correlation which would be found between two similar forms of the tests were a second form available. It is no doubt true that certain of the patterns are more reliable than others, a fact which militates against the reliability of the scale as a whole.

(2) The validity of certain of the individual tests is doubtful. The inter-correlations in patterns II and III are low, pattern I being the most satisfactory. The same situation holds for the correlations of the individual tests with the composites of the remaining tests of the same pattern.

(3) The lack of inter-correlation between tests of the same pattern in the case of pattern II, and to a less degree in the other patterns, may be due to either: (a) unreliability of the individual tests (not determined in this study), or (b) faulty location of certain tests. In the latter case further study might suggest changes from one pattern into another.

(4) There was found no reason to believe that the Downey tests measure merely intelligence. The inter-correlations, where they are significally high, are not due to the common factor of intelligence.

(5) The attempt to validate the separate tests of the scale by means of teachers' judgments yielded ambiguous results. It would appear that strict validation of the Downey tests by the method of estimates would be a very difficult task. We can conclude that, with a few exceptions, high school teachers cannot identify the psychological traits measured by the Downey tests by study of the characterizations of those traits which have been published by the author of the tests. It should be pointed out that the estimates used in this study cannot be assumed to be valid, although their general reliability is doubtless satisfactorily high for purposes of a rough criterion. The question of the validation of the Downey tests can better be dismissed by refering the reader to Dr. Downey's manual, previously cited, for a discussion of this point.

(6). In general there are a considerable number of tests in the *Downey Will-Temperament Test* that show promise of lending themselves to further study and development. Duplication of such tests and probably a number of substitutions will lead to a more valid instrument for the measurement of volitional traits. The urgent need for such measurements demands that the pioneer efforts of Dr. Downey be rewarded with the recognition that much promising ground had been broken by her patient investigations and that further progress has been made possible in this direction.

THE FRESHMAN: THORNDIKE COLLEGE ENTRANCE TESTS, FIRST SEMESTER GRADES, BINET TESTS

By W. T. Root,
University of Pittsburgh.

INTRODUCTION. At the beginning of the academic year, 1921-22, the College of the University of Pittsburgh* gave the Thorndike Intelligence Examination for High School Graduates to a Freshman class of 569. The tests given were: Part I, Form I (designated as T1); Part I, Form M (T2); Part II, Form D (T3); Part III, Form D (T4). The total, that is the sum of the four tests, is designated T5. The test was given without intermission other than the short intervals provided for in the instructions.

Results. The following averages were secured:

Test	Average
T1	105.119
T2	105.276
T3	101.504
T4	32.97
T5	341.61

Final score (T5 divided by 5) 68.38
Median score, 70
Range 21 to 110

Correlations between different forms. The forms were correlated with the total singly and in combination:

Test	r (Pearson)	P. E.
T5 with T1	.848	.007
T5 with T2	.829	.009
T5 with T3	.904	.005
T5 with T4	.754	.01
T5 with T1 plus T2	.914	.004
T5 with T1 plus T3	.960	.002

---

*The Department of Psychology gave and scored the tests. The writer is indebted to the graduate students in psychology for many weary hours of statistical calculation. Special acknowledgment is due Professor Florence M. Teagarden for transmuting grades into sigma values and for giving the Binet Tests to the Freshmen girls. Mrs. Ida Nyce Root is to be thanked for many days spent in the vexatious task of statistical rechecking and of ferreting out errors.

The high correlation of T1 plus T3 with T5 suggests very strongly that T2 and T4 be eliminated in future testing. This would mean a great economy in the time required to give the test, in the time required to score, and in the cost of test materials.

Correlation of T5 with the academic grades for the first semester. In attempting to summate the academic grades the writer met with the usual difficulties. It seemed best after all factors were considered to resort to the tedious method of transmuting all grades to sigma values thus correcting for the great variance in distribution of A's, B's, C's, D's, and F's. The sigma value was weighted according to the number of units of credit the course in question carried. The weighted sigma values were added and divided by the number of units of credit carried by the particular Freshman in question. This procedure overcomes the injustice usually experienced by students carrying "stiff" courses (i.e., courses with a high percentage of D's and F's and a low percentage of A's and B's). The sigma values of A, B, C, D, and F were determined for each subject taught to Freshmen. D is the lowest passing grade while F is failure.

The correlations between the academic grades and the Thorndike forms are as follows:

Test		r	P.E.
T1	with grades	.42	.02
T2	with grades	.33	.03
T3	with grades	.36	.03
T4	with grades	.415	.02
T5	with grades	.51	.02

The Thorndike total was now correlated with each college subject taken by the Freshmen, debarring a few in which the number of beginning students was too small to admit of correlation. (Tabulation I)

Academic College Subjects: Grade Percentages, Sigma Values and Correlations With Thorndike College Entrance Tests.

College Subjects	Units of Credit	GRADES															Correlations between Academic Grades and T. College Entrance Tests		
		A			B			C			D			F					
		Percentage of	σ Value	Weighted σ Value	Percentage of	σ Value	Weighted σ Value	Percentage of	σ Value	Weighted σ Value	Percentage of	σ Value	Weighted σ Value	Percentage of	σ Value	Weighted σ Value	Number of Cases	r	P. E. of r
Biology 1	4	12.16	4.54	18.12	31.08	3.59	14.36	28.37	2.81	11.24	20.27	2.10	8.40	8.10	1.27	5.08	73	.495	.06
Biology 51	4	6.25	4.84	19.36	21.25	3.95	16.80	30.00	3.19	12.76	27.50	2.44	9.76	15.00	1.57	6.28	151	.525	.05
Chemistry 1	4	4.00	5.03	20.12	32.00	3.80	15.02	26.00	3.00	12.00	16.00	2.45	9.80	21.00	1.75	7.00	258	.434	.03
English 1	3	6.00	4.86	14.58	27.00	3.86	11.58	50.00	2.81	8.43	15.00	1.70	5.10	2.00	.73	2.19	495	.358	.08
French 1	3	13.00	4.50	13.50	23.00	3.69	11.07	31.00	2.97	8.91	21.00	2.25	6.75	12.00	1.46	4.38	72	.396	.07
French 3	3	13.00	4.50	13.50	32.00	3.55	10.65	37.00	2.66	7.98	14.00	1.78	5.34	4.00	.97	2.91	131	.453	.05
French 7	3	21.00	4.19	12.57	43.00	3.16	9.48	14.00	2.42	7.26	14.00	1.93	5.79	7.00	1.21	3.63			
German 1	3	10.00	4.63	13.89	25.00	3.75	11.25	30.00	3.00	9.00	20.00	2.33	6.99	15.00	1.57	4.71	60	.503	.06
Graphics 2	2	3.00	5.14	10.28	8.00	4.47	8.94	40.00	3.50	7.00	43.00	2.47	4.94	6.00	1.14	2.28	109	.348	.06
History 1	3	11.00	4.59	13.77	25.00	3.72	11.16	47.00	2.76	8.28	16.00	1.67	5.01	1.00	.51	1.53	122	.483	.05
History 7	3	9.00	4.66	14.04	22.00	3.84	11.52	44.00	2.93	8.79	14.00	2.09	6.27	11.00	1.41	4.23	48	.429	.08
History of Religion	2	9.00	4.68	9.36	32.00	3.67	7.34	41.00	2.72	5.44	18.00	1.67	3.34	0.00	0.00	0.00			
Human Prog.[1]	2	4.00	5.03	10.06	29.00	3.89	7.78	45.00	2.87	5.74	19.00	1.86	3.72	3.00	.86	1.72	527	.689	.01
Latin 5	3	21.00	4.25	12.75	42.00	3.20	9.60	25.00	2.31	6.93	8.00	1.60	4.80	4.00	.97	2.91			
Math. A	3	5.71	4.88	14.64	22.86	3.94	11.82	22.86	3.25	9.75	31.43	2.56	7.68	17.14	1.64	4.92	33	.578	.08
Math. 1	3	25.49	4.13	12.39	13.72	3.45	10.35	21.57	3.00	9.00	25.49	2.38	7.14	13.73	1.52	4.56	52	.512	.07
Math. 2	3	12.50	4.52	13.56	12.50	3.88	11.64	33.33	3.21	9.63	29.16	2.39	7.17	12.50	1.48	4.44	23	.614	.09
Math. 21	4	6.31	4.84	19.36	15.32	4.08	16.32	25.22	3.40	14.60	32.43	2.67	10.68	20.72	1.75	7.00	110	.390	.06
Physics 1	4	10.00	4.63	18.52	17.00	3.89	15.56	31.00	3.19	12.76	21.00	2.52	10.08	21.00	1.75	7.00	134	.498	.04
Psychology 3	3	18.00	4.33	12.99	9.00	3.75	11.25	18.00	3.35	10.05	46.00	2.54	7.62	9.00	1.32	3.96			
Pub. Speaking 1	2	10.61	4.06	8.12	42.47	2.95	5.92	42.47	1.86	3.72	3.53	.65	1.30	.87	19.00	.38	108	.458	.05
Spanish 1	3	12.00	4.54	13.62	19.00	3.78	11.34	32.00	3.07	9.21	20.00	2.39	7.17	17.00	1.64	4.92	71	.472	.06
Spanish 3	3	19.00	4.30	12.90	22.00	3.52	10.56	25.00	2.92	8.76	16.00	2.36	7.08	18.00	1.67	5.01	14	.666	.08

[1] Human Progress is a required Freshman Course similar to the course on Contemporary Civilization offered at Columbia University.

The correlations are moderately high and considerable significance as N is large in the majority of cases and the group is homogeneous. The correlation of the Thorndike test with Human Progress is conspicuously high (r. .689) with a very small probable error (P. E. .01). This course deals with a large number of the arts and sciences in a semi-philosophical way. The classes are large and a premium is placed on general information and linguistic ability. English 1, Graphics 2, Mathematics 21 and French 1 correlated less highly than the other courses with the Thorndike total. English may be a matter of home conditions, foreign parentage or drill and follow up work through four years of high school. The students enter Graphics and Mathematics with very different preparations; French may be very easy for the student with a fair knowledge of Latin or one of the Latin languages. If Italian is spoken in the home it may prove to be a hinderance in an English course and a boon in a French course. One may speculate indefinitely as to the general causes for correlation or non-correlation.

The prognostic value of the Thorndike College Entrance Tests. It is possible to have reasonably high correlations without much justification for applying the findings to an individual case. Distinct sets of causes may be at work: thus, we may have nearly perfect correlation with a portion of the cases, a moderately positive trend for a large median group, with a significant group showing a negative correlation. Again, we may have marked negative or positive correlation in certain regions of the data with an indifferent correspondence in other regions. It becomes pertinent, then, to ask for an *analysis* of the tendency to correlation or rank-correspondence as well as to compute the coefficient of correlation.

The chief reason for administering the tests now under consideration to Freshmen is to predict their capabilities in connection with academic college requirements. It therefore seemed desirable to rate the Freshmen in some form of grade classification, arranged in rank-order, and to compare this ranking with the Thorndike rank-orders. The conventional distribution of grades over a $\pm 3\sigma$ normal distribution curve is, perhaps, as good as any. The data can then be tabulated as follows (A, B, C, D and F referring to grades and α, $\beta\gamma$ and ϵ referring to precisely the same percental divisions in the descending rank-order arrangement of the Thorndike totals):

	A or α	B or β	C or γ	D or δ	F or ϵ
%	3.5	24	45	24	3.5
No. cases........	20	137	255	137	20

The college grades are expressed as sigma values and each Freshman's rank-order is determined by finding his average sigma (the sum of his weighted sigmas divided by the number of units carried). These average sigma values arranged in descending rank-order are then ready for comparison with the corresponding rank-orders in the Thorndike tests. If the Thorndike test is prognostic of academic ability then the 20 students composing the A group should have a rank-correspondence such as to place the majority of them in the α group (the 20 ranking highest in the Thorndike group). Tabulation II presents the rank-orders for college grades and Thorndike scores for comparison and analysis of the rank-correspondence in different percental regions of the rank-order series. The question is: Does the Thorndike Test performance anticipate the quality or approximate quality of college work a particular Freshman will do?

TABULATION II.

	A	B	C	D	F	TOTALS
A	4 1	9 5	7 9	0 5	0 1	20
B	7 5	60 33	54 61	11 33	5 5	137
Γ	8 9	62 61	115 115	65 61	5 9	255
Δ	1 5	6 33	70 61	53 33	7 5	137
E	0 1	0 5	9 9	8 5	3 1	20
TOTALS	20	137	255	137	20	569

The following observations may be made directly from the tabulated data:

1. No. α's are ranked D or F in academic standing.

2. No. ϵ's attain A or B standing.

3. But of those making A standing, only 1/5 are α's and the largest number are drawn from the γ group.

4. Had superior scholarship been predicted by selecting the α group we would actually have secured 4 A students along with 9 B students and 7 C students.

5. 80 of the 157 students comprising the A and B groups are in the α and β groups.

Seventy-one of the 157 comprising the D and F groups are in the δ and ϵ groups.

That is, the upper 27.5% of the Thorndike rank-orders would have been somewhat selective as regards academic attainment. Had the ϵ group been taken as a basis of entrance exclusion, it would have resulted in the excluding of 8 D students and 9 C students out of a total of 20 ϵ students.

6. Had the college wished to maintain very high standards and had excluded δ and ϵ standings, it would have excluded 79 C students, 6 B students, and 1 A student out of a total of 157 Freshmen. The Thorndike ranking would have correctly placed 71 (that is δ-ϵ's who are in corresponding academic rank-orders, namely the D-F groups), while random choice alone would have correctly placed 44 of these.

7. On the other hand, if the α-β-γ groups had been used as a basis of initial retention in college, 86 D and F students would have been admitted into the Freshman class against the 71 D and F students excluded in the δ-ϵ's exclusion. Chance alone would have given us 114 D and F grade Freshmen, out of a total of 410 α-β-γ's.

Of the 569 students chance alone would have given 157 DF's. Had the δ-ϵ groups been arbitrarily excluded (and the α-β-γ groups retained), 71 D-F students would have been excluded and 86 D-F included. But 86 A-B-C students would have been thus excluded. This would give a total of 176 improper inclusions and improper exclusions or 19 more than the improper inclusions chance alone would have given.

8. There are five comparable groups, α-A, β-B, γ-C, δ-D, and ϵ-F. Of these groups, the γ-C is the only one that is largest for the column in question. In this case it is totally without prognostic significance as chance alone would have placed 115 C's in the γ group. It should furthermore be noted that the distribution of C students through α, β, δ, γ and ϵ is about the distribution

chance would have given. And likewise the row has a close-to-chance distribution through A, B, C, D and F. Hence for a student to be in the γ group argues nothing as to his academic rank-order; and vice versa, C grades run the gamut of a near-chance distribution curve from α to ε.

Deducting from each of the five comparable groups (A-α, B-β, etc.) the number of cases that would fall in these groups by chance alone, gives the number of cases precisely located in a *corresponding* rank-order. We have:

Of 20 α's, 3 A's or 15% in precise corresponding rank-order.
Of 137 β's, 27 B's or 19% in precise corresponding rank-order.
Of 255 γ's, none.
Of 137 δ's, 20 or 15% in precise corresponding rank-order.
Of 20 ε's, 2 or 10% *in precise corresponding rank-order.*

9. Compiling β, γ and δ we have 529 cases, comprising 128 B's, 246 C's, and 129 D's. Thus the Thorndike tests would have correctly placed 12 more subjects than chance alone or about 2 per cent of the total number of students in the combined β, γ and δ groups.

TABULATION III.

A	Scores over 95	9	2	0	3	5	1	0	0
B	Scores 85 to 95	55	10	2	17	24	7	1	4
C	Scores 70 to 85	203	36	6	25	92	45	23	12
D	Scores 60 to 70	143	25	0	8	53	37	27	18
E	Scores 50 to 60	96	17	1	2	18	27	34	14
F	Scores under 50	63	11	0	0	11	26	11	15

AN ANALYSIS OF THORNDIKE'S GROUPINGS

Thorndike attempts to prophecy on a basis of the absolute scores received. He makes the following classification:

"A boy scoring over 95 is worth admitting in almost entire disregard of technical deficiencies.
A boy scoring 85 to 95 has intellect enough to do collegiate and professional work with distinction.
A boy scoring 70 to 85 has intellect enough to do the work to obtain a college degree.

A boy scoring 60 to 70 may be admitted if he is sufficiently in earnest and otherwise desirable.

A boy scoring 50 to 60 should be admitted only if he is of extraordinary zeal or has suffered very great educational handicaps.

A boy scoring under 50 should not be admitted."

The accompanying tabulation (III) follows Thorndike's grouping. The academic grade rank-orders have been arbitrarily classified to agree in rank correspondence with the Thorndike classification.

Conclusions:

1. None of the Thorndike A group falls in the corresponding academic group a (the 9 highest academic rank-orders); however, none falls in the lowest academic groups, e and f.

2. Likewise none of the F group falls in the academic a or b groups.

3. The students composing the a group came chiefly from the C group.

4. Taking Thorndike's C group as the dividing line for inclusion in or exclusion from college entrance, one would include 174 correctly but also 93 incorrectly, (students making d, f, and e grades). On the other hand, the elimination of the D, E, and F students would have excluded 93 students who are doing work in the a, b and c groups. Thus there would be 186 improper inclusions and exclusions in using Thorndike's A, B and C groups as an inclusion basis against 209 proper exclusions. Out of 569 students 302 would be excluded. It must be recalled that in excluding such a large percentage of the students seeking admission, chance alone would have correctly excluded 161.

5. Excluding E's and F's would retain 85 of the upper Thorndike groups doing no better work than E's and F's, and would exclude 85 E's and F's doing work on a par with those admitted. Granting that e and f grade work warrants exclusion, we have 74 proper exclusions, and 170 improper inclusions and unjust exclusions.

6. Let us take Thorndike's lowest group, F, (scores below 50), which Thorndike categorically asserts should be excluded. Of the 63 comprising this group 48 do work on a par with many in the retained groups, while 48 in the retained groups are doing the same quality of academic work as those excluded. Thus we have 96 improper exclusions and inclusions, and 15 proper exclusions.

7. Tabulations II and III lead to the same general conclusions. *While the correlation between grades and Thorndike scores is reasonably high, an inspection of the differences in rank-order*

correspondence indicates that great injustice would be done, both in inclusions and exclusions, were the entrance tests used as a basis of college admission. In actual practice the lines of demarcation between groups are too illy defined to warrant such drastic action as off-hand exclusion from college. However, the trend of correlation is sufficiently strong to far surpass any impressionistic estimate and the tests should constitute an excellent basis for the initial analysis of cases, and for judicious use in attempting to size up Freshmen failures. Whether the mass test is more truly prophetic of final college success than the combined estimates of dozens of instructors after seventeen weeks contact is a speculative matter which can be decided only by follow-up work for a number of years.

EXTREME UPPER AND LOWER RANK-ORDERS IN COLLEGE ENTRANCE TESTS COMPARED AS TO PERFORMANCE IN BINET TESTS

Forty of the upper rank-orders of the Thorndike Test (1to 58 inclusive) and also 40 of the lower rank-orders (479 to 569 inclusive)[1] were given the Binet Tests.[2] The accompanying tabulation (III) shows clearly that the correspondence between the two tests is extremely close, quite in contrast to the vague correspondence between the Thorndike and the academic standing.

---

[1] Inability to arrange a time suitable for both the student and the examiner giving the Binet Test, made it necessary to go below rank 40 and above rank 520 in the lower group. One of the lower rank-orders was eliminated. The subject, an intelligent Italian just learning the language, refused to try the Binet Test after doing the vocabulary.

[2] The student appointments for the Binet Tests were arranged through the co-operation of Professor B. F. Ashe, Freshman Advisor.

TABULATION IV

Binet Tests Used as a Basis for Comparing the Extreme Upper and Lower Rank Orders of Thorndike College Entrance Tests

		40 highest Rank Orders	39 lowest Rank Orders
Chronological age range		22-8 to 16-5	27 to 18-2
Chronological age median		20-1	18-5
Mental age range		19-6 to 17-2	18 to 12-1
Mental age median		19	15-5
Vocabulary range		16,560 to 12,780	12,600 to 6,840
Vocabulary median		14,940	10,260
Number failing			
XVIII	Superior adult vocabulary	5	39
	Binet's paper-cutting test		25
	Repeats 8 digits	13	24
	Repeats thought of passage		26
	Repeats 7 digits backward	12	31
	Ingenuity test	8	31
XVI	Average adult vocabulary		30
	Interpretation of fables		15
	Difference between abstract words		11
	Problem of the enclosed boxes.	5	22
	Repeats 6 digits backwards	1	17
	Code		23
XIV	Vocabulary 50 words		8
	Induction test		7
	President and king		8
	Problems of fact		10
	Arithmetical reasoning		9
	Clock		8
XII	Vocabulary 40 words		1
	Abstract words		
	Ball and field		4
	Dissected sentences		3
	Fables		1
	Repeats 5 digits backwards		3
	Pictures interpretation		1
	Gives similarities		1
X	Vocabulary 30 words		
	Absurdities		1
	60 words		2
Total No. of tests failed		44	361

The lower group is made up of older students and it may be added has more of foreign birth and foreign family. It is a matter of general observation that older persons show less associative plasticity and speed.

Concerning I. Q. and mental age:

8 of the lower group had I. Q.'s 100 or more.
31 of the lower group had I. Q.'s less than 100.
3 of the lower group had mental ages of 12(11-7 to 12-6)
5 of the lower group had mental ages of 13(12-7 to 13-6)
4 of the lower group had mental ages of 14(13-7 to 14-6)
8 of the lower group had mental ages of 15(14-7 to 15-6)
12 of the lower group had mental ages of 16(15-7 to 16-6)
4 of the lower group had mental ages of 17(16-7 to 17-6)
2 of the lower group had mental ages of 18(17-7 to 18-6)

In the upper group the lowest I. Q. was 104. Nineteen failed in the high adult digits tests; 7 failed in both, 14 passed all of the tests; 13 failed 1 test; 11 failed 2 tests; 1 failed 3 tests; and 1 failed 5 tests.

The fact that 25 of the 44 failures in the upper group were digit tests indicates that the digit tests (counting an entire year of mental age) are entirely out of proportion to their importance in determining mental age. A careful study of the individual cases gives little evidence that immediate verbal auditory memory span for digits is of any marked value in determining superior intelligence. It is a matter in which there is marked individual difference in adults with slight reason for believing that there is a corresponding intellectual difference.

Reading and Vocabulary. The most striking contrasts between the groups is in the vocabularies and reading habits. The *highest vocabulary in the low group is less than the poorest vocabulary of the high group.*

An attempt was made to get at the reading habits of the groups under consideration. A rough reading scale was improvised, scoring as follows:

1. Readers of critical and difficult literature.
2. Superior readers, read classical and good current literature.
3. Average readers.
4. Readers of newspapers and sport news.
5. Readers who do almost no voluntary reading being nearly devoid of a reading habit.

Tabulating for the two groups we have:

Reading Rank	Number in High Group	Number in Low Group
1	6	1
2	12	2
3	15	7
4	7	14
5	0	15

The contrast here is very marked indicating decided lack of reading habits in the lower group.

TABULATION V.

Cases With Marked Inconsistency in Performance in Upper and Lower Groups

Thorndike Test Rank	Academic Rank	Reading Rank	Voc. Score	Mental Age Binet	REMARKS
3	386½	3	91	19	Out of school 3 yrs. Trig. & Physics difficult
5	132½	2	92	19	
6	137	3	92	19	
8	188½	1	86	18½	Out of school 3 yrs. Indifferent to grades
10	197½	2	83	19½	Exemption grades in H. S.
13.5	188½	3	81.5	19	
20	297	1	87	19½	
27½	223½	2	79	19	
30	197	1	90½	19	
35½	554	2	88	19	Absent one month in first semester
35½	301½	4	84	19½	
40½	542	2	81	18½	
48½	350½	3	73	18½	
58½	258	4	89.5	17-2	
522	283	3	58	13-6	Italian. Foreign parent.
524	71	2	56	16-5	Jewish. Soho district.
526	172	2	67	15-8	Jewish. Foreign parent.
506½	290½	3	70	17-1	Attended sixteen different schools.
538½	249	4	62	16	
545½	267½	5	40	14-9	Italian. Foreign home.
553½	264	5	63	13-11	Jewish.
554	208½	4	70	17-7	Jewish.
557	157½	5	48	12-2	Negro.
561	206	4	68	18	Commutes four hours daily.

Analysis of cases showing marked contrasts in performance. It would have been highly desirable to have studied the social conditions surrounding these Freshmen composing the upper and lower groups. This was impossible, due both to lack of time and to the diffidence on the part of students concerning their social and economic status. Nevertheless, after personal interviews with the students in question, it is possible to throw some light on the marked inconsistencies in test performance and academic rank, presented in the accompanying Tabulation V.

CONCLUSIONS

1. A high Thorndike rank is accompanied by a high Binet score and large vocabulary; and vice versa, a low Thorndike score by low mental age and small vocabulary.

2. But high vocabularies are not necessarily found with a superior reading ranking or vice versa.

3. Of the 40 in the upper Thorndike group, 14 are distinctly below any corresponding academic rank-order, while 7 fall below the median rank-order, 285, of the entire Freshman class.

4. In the lower group just the opposite is found; with low mental age, low vocabulary and low entrance examination score is found a relatively high academic rank-order. Of the 10 cases, 9 have academic ranks above the median rank-order of the Thorndike scores.

5. The cases showing marked inconsistencies constitute nearly 1-3 of the total number of the 79 cases studied intensively.

6. It should also be noted that in spite of exceedingly low vocabulary and mental age, cases 522, 553½, and 557 make more than median grades.

7. In the lower group, 6 are either Italian or Jewish. In all of these cases, and many more not presented here, the parents speak a foreign language. Is it not to be expected that an English vocabulary and tests resting on the niceties of English idiom will suffer?

8. Absence during the semester and a several-year gap between high school and college seem to be the most obvious and plausible explanations in certain cases of relative academic inferiority in the upper group.

9. The negro, case 557, unlike the Jewish and Italian cases, appears very dull. He is carrying, however, a very heavy program including mathematics, physics and chemistry in which he is making more than median grades. An examination of the percentage of A's, B's, C's, D's and F's given in these courses indicates that they are not "pipe courses," (see Tabulation I). In the opinion of the writer the explanation is this: The student has

had an excellent high school training with a careful drilling in technical courses that dovetail into the courses taken in college. It is doubtful if the chap will long maintain an academic ranking of 157½ against a Thorndike rank of 557 and a Binet mental age of 12-2.

10. The personal interviews elicited the following, at least plausible, explanations for the marked rank-order difference between academic grades and the Thorndike tests:

(a) Some students are capable but indifferent to grades, tending to follow an interest rather than make punctilious preparation of daily tasks. Case 8 is a good example. In the opinion of the writer, the Thorndike test places him more nearly correctly. The chap has been out of school three years between high school and college, which is perhaps a contributing factor, but the chief difficulty is that he has very positive ideas concerning academic values and refuses to fulfill, what he considers to be, petty classroom demands. He has already been successful in the work-a-day world.

(b) Outside work, undoubtedly, in many cases makes for poorer grades.

(c) Commuting tends, also, to produce a poorer grade of work than otherwise.

(d) The character of the high school work plays a large part in the character of the Freshman college work. For example, one high school, gives to its seniors a course on Current Literature. All of these students show a relatively greater general information than the average high school graduate. Another high school apparently gives unusual work in science and mathematics. One of the students from the lower Thorndike group, with good grades, stated that the physics and mathematics he was taking were almost entirely review. Again, a student changing from a literary or a scientific course in high school to a scientific or literary in college is greatly handicapped over the one who has had a constant vocational objective. All considered, students who have established good habits in mathematics, science, laboratory and English composition, regardless of their native endowment, have an advantage over those who have not.

(e) Unusual zeal for learning and persistence mark the attitude of many of the foreigners, especially the Jews. The greater the economic and cultural handicap, seemingly, the greater the zeal.

(f) One might elaborate indefinitely, but this is sufficient to indicate that a careful study of the individual case is the next step after the rough grouping of a mass intelligence test. It is much easier, and far more satisfying, to deal in broad statistical trends than it is to tease out the vexatious problems of individual causes in variation. —— But, the latter is the next logical step in testing.

11. *Errors in grading undoubtedly play a part in the rank-order* differences between the Thorndike rating and the instructors rating. Impressionistic methods of grading, marked bias in weighting certain qualities in work and too few facts for estimate often cause the instructor's grade to be a long way from the absolute facts.

CONCLUSIONS

1. *While there is a fairly high correlation between academic* grades and the Thorndike tests, upon analysis the approximation is found to be too rough to permit exclusion, inclusion, or the assumption that the highest test group (above 95) is peculiarly superior.

2. Nevertheless, the college entrance tests are not without merit, offering as they do, an initial point for analysis and judicious prognosis, but unfortunately, not scientific prophecy. Such tests are undoubtedly superior to any impressionistic estimates of the Freshman. It becomes necessary, however, to follow up the test findings with a careful analysis of the qualitative character of the score; to supplement with a common-sense investigation of the nature of the high school training noting how well it dovetails into the curriculum of the particular student in question. The home, language, zeal, outside employment, and so on must be weighed and evaluated. The tests form but a part of a complex differential diagnosis. All factors must be carefully summated analogous to a court's getting a perspective on circumstantial evidence.

3. Whether the student is surpassing or falling below the Thorndike rank-order, he is equally of interest. At present such cases demand an investigation to "show cause."

4. Not only is the Freshman up for critical study and evaluation but the test purporting to evaluate him is equally on trial to show it has value.

5. A combination of Part I, form 5, and Part III, form D would probably be as valuable as all four forms; thus enacting a saving of 50 per cent in giving, scoring, and purchase price of blanks. But it should be added that for purposes of analysis, much is to be gleaned by a study of the qualitative character of all four forms.

6. The rank-order comparison between the Thorndike tests and the academic grades should be continued throughout the college course. At present any assumption as to the superior prophetic value of the mass intelligence test over the Freshman's academic performance, is speculative and idle.

THE UNRELIABILITY OF M. A. AND I. Q. BASED ON GROUP TESTS OF GENERAL MENTAL ABILITY.[1]

By ARTHUR I. GATES
Teachers College, Columbia University

The practice of computing mental ages and Intelligence Quotients from scores of group tests of general mental ability has been questioned recently on theoretical grounds,[2] and various workers are discovering the unreliability of these measures by comparing the results obtained from different tests.[3] This paper presents data which show that at least under certain conditions, quite ridiculous IQ's may be derived from certain group tests.

The following tests were given on successive days to the pupils of Grades I to VIII inclusive of the Scarborough School: Dearborn, Examination I or II, Haggerty, Delta I or II, Holleys Picture Completion or Sentence Completion, Illinois (III to VIII only), Otts Primary or Advanced, National Intelligence, Forms A and B, (III to VIII only) and Meyers Mental Measure. The Stanford-Binet was given during the year to Grades I to VI inclusive.

The tests were given to groups of twenty or less and particular attention was given to time, instructions, provision with pencils, etc. The pupils had learned to take tests earnestly and we are confident that the test conditions in general were more uniform and satisfactory than will be obtained in ordinary school practice. The results were scored by an expert.

The mental age equivalents of the scores on the several tests were computed by use of the age norm tables accompanying each test.[4] In most cases the line describing the median age score was extended to cover such of our scores as exceeded the highest age norm given. From these data, IQ's were computed.

Table I, which gives the data in detail for Grade I, illustrates the treatment of results. The horizontal lines give the facts for the several subjects; the vertical columns give the facts for the class. Columns 1 to 6 give the IQ's which are summarized at the foot in the form, (1) the highest IQ in the class (2) the lowest, (3) the mean and (4) the mean deviation from the mean. Column 7 gives the mean IQ for each subject; column 8 gives the mean deviation of the subject's various IQ's from his mean and column 9 gives the spread i. e., the subject's highest IQ minus his lowest IQ. These data are summarized at the foot of each column.

---

[1] This study is a contribution of the Department of Educational Research, Scarborough School, Scarborough, New York.

[2] F. N. Freeman. The Interpretation and Application of the Intelligence Quotient. Journal Educational Psychology. January, 1921. L. M. Terman. Mental Growth and IQ. Journal Educational Psychology. October, 1921, especially pp. 398-401.

[3] For example, J. L. Stenquist, Journal of Educational Research, Dec., 1921.

[4] For the N. I. T., Termans Vallejo norms, published in the Journal of Educational Research, February, 1921, were used.

The facts of Table I may be considered with reference to certain important school problems.

1. Is the class, on the whole, composed of bright or average pupils? A very general answer — bright — may be given, but the degree of brightness varies widely according to different tests. The mean IQ's range from 109 (Dearborn Examination) to 129 (Myers). According to the one the group should do slightly better than average work, according to the other, very much better. According to the one the average pupil could scarcely complete a college course, according to the other he could do so with ease. Assuming the average chronological age to be six years according to one test the mental age would be 6.5 years corresponding to first grade ability, but according to the other test the mental age would be 7.8 years corresponding to advanced Grade II work.

2. Are the pupils of uniform or diverse brightness? The figures for the lowest IQ are fairly uniform, ranging from 92 to 99 but those for the highest are diverse, ranging from 121 to 165. The mean deviation of IQ's from the average are for the Dearborn 6.0 and for the Myers 12.3, with figures for the other tests scattering between these extremes. Recommendations concerning sectioning for purposes of promotion would depend very greatly on what test is used.

3. What is the IQ of a particular pupil? Pupil A is given IQ's ranging from 104 to 165, that is, by one test he would be classified as "average" by another a "genius." Pupil A has various IQ's ranging from 104 to 144. Column 9 gives the difference in points of IQ between the highest and lowest. The largest difference is 61, the smallest 15, the mean 32. On the average a pupil ranges from "average" to "very superior" intelligence depending on the test. If we take as measures of variability the mean deviation from the average IQ instead of the whole range the figures given in column 9 are of course smaller the greatest being 17.0 the smallest 3.8, the mean 9.3. That is if the average IQ is 118 in the average case half of the IQ's obtained from the several tests will fall between the limits 108.7 and 127.3. Further statistical treatment is unnecessary. It is obvious that the variability of the IQ's obtained by several tests all given within a few days is so great as to make prediction concerning optimum rate

TABLE I

SHOWING I .Q.'S OBTAINED BY PUPILS OF GRADE I IN SEVERAL TESTS

	1	2	3	4	5	6	7	8	9
	Stanford Binet	Dearborn Exam. I.	Otis Prim.	Kingsbury	Haggerty Delta I.	Myers M. M.	Mean IQ	Mean Dev. from Mean	Range
A	124.	104	135	124	125	144	120	9.0	40
C	120.5	121	144	114	125	131	126	7.7	30
D	114.	107	122	134	102	122	117	9.2	32
E	135.	115	138	126	108	121	124	9.2	30
F	110.	117	110	111	112	129	116	5.3	19
G	133.	112	132	116	112	125	122	8.3	21
H	92.5	99	119	105	120	115	109	9.5	28
I	104.	107	140	116	111	135	119	12.5	36
J	96.	99	117	104	107	152	113	14.8	56
K	103.	92	98	98	107	99	100	3.8	15
L	120.	115	131	127	114	136	124	7.5	22
M	108.	108	125	100	99	123	111	9.2	26
N	110.	111	135	118	125	114	119	7.5	25
O	104.	109	139	117	106	165	123	17.0	61
P	121.	115	137	98	117.5	117.5	118	7.7	39
Highest	135.	121	144	134	125	165	126	17.0	61
Lowest	92.5	92	98	98	99	99	100	3.8	15
Mean	113.	109	129	114	113	129	118	9.2	32
M. D.	10.1	6.0	7.6	8.9	6.8	12.3	5.4	2.2	9.6

SUMMARY OF RESULTS FOR GRADE II n=10.

Highest	141	147	148	155	150	196	150	17.5	76
Lowest	103	97	122	108	92	105	114	5.3	21
Mean	117	116	132	135	124	135	127	10.0	35
M. D.	7.4	7.3	6.9	12.4	12.6	15.4	7.6	2.8	10

of promotion, possibility of completing college, etc. quite out of the question.

RESULTS FOR GRADE II

The results for Grade II are summarized in the lowest section of Table I. For this grade, the group tests, all of them non-verbal, are the same as those used in Grade I.

The results are similar. The IQ's yielded by the Myers test show an average deviation from the mean which is more than twice that given by the Otis. One individual receives an IQ on one test which is 76 points higher than that given by another test; the difference between an average intelligence and idiocy. The average range of IQ for an individual is 35 points.

RESULTS FOR GRADES III to VIII

Table II gives summaries of the data for grades III to VIII. This Table shows the ridiculous results which teachers and other workers are likely to obtain from efforts to derive M. A.'s and IQ's from group tests in these grades. For example, one fairly bright boy in Grade VII earns a mental age of 46 years on one test giving an IQ of 408. If they should take the results of certain tests seriously there are many pupils who could think rather pitifully of Sir Francis Galton whose IQ is estimated by Terman to be only about 200. According to the Myers test the average IQ's of our Grades VII and VIII is 261.[5] Picture the perplexity of the teacher test enthusiast who finds that pupil K in Grade V earns an IQ of 99 on the National test and 366 on the Myers. Many others are nearly as baffling.

The variability of scores obtained by a given individual becomes greater as the grade becomes higher, as may be seen in Table III which gives the range of IQ's and M. D.'s for the average individual in each grade.

These facts should not be interpreted as a sweeping criticism of the usefulness and validity of group tests of general mental ability. What is urged is a restraint of certain inappropriate uses of these instruments. We should not, it is clear, compute M. A.'s and IQ's from raw scores, but these raw scores themselves may be put to profitable use. It has been shown elsewhere that these group tests at least those very verbal in content give substantial correlations with achievement in school work.[6]

---

[5] IQ's could not be computed for National and Otis scores for most of our Grade VII and VIII pupils because they earned scores considerably higher than those represented by the line when it becomes horizontal.

[6] A. I. Gates, Correlations of Intelligence tests with achievement in school subjects. Journal of Educational Psychology, March, April and May, 1022.

TABLE II

SHOWING HIGHEST, LOWEST, MEAN AND M. D. OF I. Q.'S FOR THE GRADE ACCORDING TO THE SEVERAL TESTS. IN LINES 1 AND 2, DIFFERENT INDIVIDUALS MAY BE REPRESENTED BY EACH FIGURE

	1 Stanford	2 N. I. T.	3 Dearborn	4 Otis	5 Haggerty	6 Myers	7 Illinois	8 Holley Vocab.	9 Mean of 7 I. Q.'s	10 M. D. from (9)	11 Range (Highest Minus Lowest
GRADE III											
Highest IQ.	*133*	*127*	*133*	*135*	*147*	*171*	*154*	*166*	*143*	*16.8*	*76*
Lowest IQ.	95	94	90	106	100	93	80	90	98	4.6	16
Mean IQ.	112	112	109	120	120	120	116	115	116	8.8	35
M. D. from Mean	8.9	7.2	8.3	5.9	10.2	20.0	15.3	16.2	8.7	2.2	11
GRADE IV											
Highest	140	133	142	399	166	176	162	174	154	61.0	292
Lowest	99	100	104	93	113	104	100	104	113	6.0	15
Mean	120	113	124	168	135	128	126	136	132	17.2	73
M. D.	7.1	6.1	5.5	61.4	12.4	11.0	10.6	14.0	10.8	10.2	52
GRADE V											
Highest	139	131	136	169	151	324	156	147	154	55.4	266
Lowest	100	100	98	115	113	81	109	112	110	5.1	28
Mean	116	113	120	139	134	169	132	133	132	19.4	88
M. D.	9.5	7.2	10.0	13.5	9.2	79	13.9	7.3	12.3	11.7	59
GRADE VI											
Highest	160	141	170	162	164	351	191	152	180	49.3	218
Lowest	83	88	80	92	95	70	92	97	90	4.8	14
Mean	120	115	123	121	130	179	130	125	132	19.3	82
M. D.	18.8	12.9	20.2	15.9	15.5	75.6	20.6	13.0	21.8	11.9	55
GRADE VII											
Highest	Not	*	183	*	155	408	181	147	210	79.0	261
Lowest	Given		109		110	74	94	104	103	7.8	26
Mean			137		132	261	138	127	159	43.9	146
M. D.			14.2		18.0	96.2	17.3	9.2	19.6	24.4	77
GRADE VIII											
Highest	Not		148	*	138	404	168	141	186	86.8	286
Lowest	Given		105		91	88	85	96	172	5.8	20
Mean			124		116	261	124	111	147	46.9	157
M. D.			9.0		8.2	56.1	14.2	8.9	13.7	16.9	54

*Scores for nearly all pupils above the equivalent of highest ages given.

TABLE III

	Range of IQ in the Several Tests for the Average Individual	Mean Deviation from Mean IQ for the Average Individual
Grade I	32	9.2
II	35	10.0
III	35	8.8
IV	73	17.2
V	88	19.4
VI	82	19.3
VI	146	43.9
VIII	157	46.9

THE CAUSES OF THE UNRELIABILITY OF IQ'S BASED ON GROUP TESTS

An effort has been made to ascertain the causes of the extreme variations in IQ given to the same individual by different tests although we probably have been unable to discover them all or to evaluate accurately those that seem likely.

First, the distribution of native mental ability in the groups upon which the tests were standardized has probably been unequal. That this is often the case was clearly illustrated in constructing norms for the National Intelligence tests.[7] Such differences are not necessarily eliminated by mere accumulation of vast numbers of measures; a fact that is now quite generally recognized. It is probably an important source of inequality of the test norms.

Secondly, the character of school training, especially in reading, writing and arithmetic, affects the results considerably. For example, if certain grades were backward in arithmetic in comparison with pupils upon whom norms were established, they would obtain lower IQ's in those tests which contained relatively large amounts of arithmetical functions other things being equal.

Again results will differ depending on the relative amounts of non-verbal material which the particular tests contain. Two investigations have shown that in Grades III to VIII the correlations between verbal and non-verbal tests are low; sometimes negative in grade groups.[8] In our results the association between verbal and non-verbal tests becomes less close as the subjects become older. This fact is a partial explanation of the fact that the disagreement between IQ's is greater in the upper grades.

---

[7] See L. M. Terman and E. D. Whitmire. Age and Grade norms for the National Intelligence Tests. Journal of Educational Research. February, 1921, and G. M. Whipple, The National Intelligence Tests. Journal of Educational Research, June, 1921.

[8] J. P. Herring. Verbal and abstract elements in intelligence examinations. Journal Educational Psychology. December, 1921, and A. I. Gates, same journal, March, April and May, 1922.

The factor of greatest importance in our results is practice in group tests. Table II shows that Scarborough pupils earned IQ's that are obviously too high. On the whole, the more practice preceding a particular test, the higher the IQ's obtained. There is no doubt that practice results in very rapid improvement in such tests.[9] There are three ways in which practice will cause variations in IQ from test to test: First, scores will advance considerably by virtue of general adaptation to test conditions; secondly, the gain will be relatively great in the particular tests which contain large amounts of identical or equivalent exercises; thirdly, the norms for the tests will be high or low, depending on previous practice in intelligence tests as well as on general mental abilities of the subjects on whom the norms were based. Do those who construct norms for their tests actually ascertain the amount, if any, of previous experience in tests their subjects have had? Without a doubt norms established on uninitiated pupils *will differ considerably, native ability being equal, from those* secured from pupils who have been tested extensively.

Again, the cumulative procedure of establishing norms probably tends in the long run to produce norms that are too low. Testers probably succeed in controlling the time better than other conditions. In the long run numbers of students fail entirely to understand methods of procedure, skip pages, lose time through distraction, broken pencils, etc., etc., whereas the factors conspiring for higher scores are fewer. Scarborough pupils were handled in small groups. They were provided with several pencils. They had learned to do their best in all "tests." Disturbances and misunderstandings very seldom occured. These factors would tend to produce differential success depending on whether a particular test contained more numerous speed or power exercises. Well practiced pupils would probably have a considerable advantage over the inexperienced particularly in the former, because they make a quicker start and maintain a steadier performance.

Finally, most of the Scarborough pupils are of superior intelligence as evidenced by the Stanford-Binet IQ's (which average about 117[10]), by achievement in school subjects and by information concerning the social status of parents. It is possible that very bright children with a little practice may exceed average children in these tests by a larger amount than the IQ ratios would suggest. In support of this view we find that some of the very bright pupils very greatly exceed the norms for the grades to which their mental ages correspond in speed of reading. In depth or power of comprehension (Thorndike-McCall) however, their excess over the average, while marked, is not nearly so great. This would make for differential success depending on the relative amount to which achievement in each of the several tests depends upon these abilities.

---

[9] For example, see Knight Dunlap and Agnes Synder. Practice Effects in Intelligence Tests. Journal Experimental Psychology, 1020, p. 357-377.

[10] This statement is made because most of the Stanford tests were given *before the experience in group tests and because, on retesting this year, it* is found that practice in group and other tests has not caused any marked increase in IQ.

It will not be profitable to speculate further; what we need is experimentation. These experiences have been sufficient to make one very cautious in the interpretation of group test scores and particularly cautious in the use of the M. A. or I. Q.

The group tests are chiefly useful for purposes of rough classification and for very rough prediction but either should be undertaken only after inquiry into several factors. First, has this pupil ever taken such tests before? If so, which, when, and how many? We need data badly to enable us to make a rough correction in such circumstances. Secondly, what has been the subject's training in school subjects? It was found, for example, that some of our Grade III pupils made enormous improvement in most verbal group tests as the result of having improved in reading during the interval between two tests. We need information concerning the transfer of improvement in all school subjects to achievement in group tests. Finally did the subject lose any time for any reason during the test? This matter assumes great importance in some of the group tests. If all of the pupils have had no (or the same amount of) previous practice in group intelligence tests if their school training has been the same and if no accidents occur the results, at least those of verbal tests, are very useful but coarse instruments for classification.

Perhaps a change from the term "intelligence tests" to some such term as "classification tests" would have a favorable effect in reducing misunderstanding by teachers and laymen and in curbing misuse by over-zealous testers. It has been well said that the work of measurement "may soon need protection from over-extravagant hopes more than from hostile criticism."[11]

---

[11] E. L. Thorndike. Measurement in Education. The Twenty-first Yearbook of the National Society Study of Education. Part I, p. 8. 1922.

NEWS AND COMMENT

The Bureau of Vocational Guidance of Harvard University has taken over the publication of the National Vocational Guidance Association Bulletin. It is their purpose to enlarge and improve this Bulletin as support becomes assured for it.

The Journal extends a cordial welcome to this new magazine in the field of vocational guidance. Professor Frederick J. Allen will serve as Editor.

A NOTE ON POPULAR PSEUDO-PSYCHOLOGICAL BELIEFS

Kornhauser and Jackson present data in a recent issue of this journal[1] showing that actual business men do not use quack systems of judging character in their employment work. The results obtained by the use of their questionnaire are rather surprising in view of the apparently successful exploitation of "applied psychology" by various charlatans who invade our larger cities, advertise extensively in the newspapers and on the bill boards, give free lectures in the best hotels or theatres and then conduct large classes for four or six weeks at so much per head. Dr. Joseph Jastrow at Wisconsin has made a special study of these quacks and the writer believes as the result of conversations with Dr. Jastrow plus his own observations in the quack-infected Northwest that the merits of psychological research and applications are imperiled by this wide spread exploitation of pseudo-psychology.

Pseudo-phychological beliefs amounting in some cases to superstitions seem to be more common than most of us realize. The writer in conducting a summer school course in Applied Psychology was amazed to discover the range of interests expressed by these adult summer school students, but even more so, by the beliefs shown in answer to a questionnaire. The range of interests included a desire to study the problem of learning, memory, how to study, business psychology, salesmanship, influencing people, personnel work, personality, vocational guidance, abnormal psychology, psycho-analysis, spiritualism and character analysis. Below is given a summary of the beliefs expressed by these students at the beginning of the course in answer to a questionnaire.

The above table speaks for itself. If mature summer school students exhibit such a variety of beliefs what must the average business man believe? It goes without saying that such a course

---

[1] "*A note for the extent to which systems of character analysis are used in the Business World.*" Jour. Applied Psy., Vol. VI, No. 3, Sept., 1922, p. 302.

SUMMARY OF BELIEFS EXPRESSED BY APPLIED PSYCHOLOGY STUDENTS

Do you believe in:	Positive Belief	Positive Disbelief	Unknown or Uncertain
1. Astrology	2	11	7
2. Chiropractic	10	9	1
3. Fortune Telling	2	18	0
4. Graphology	5	5	10
5. Hypnotism	12	8	0
6. Memory Systems	12	7	1
7. Mental Telepathy	12	8	0
8. Spiritualism	5	15	0
9. Palmistry	1	18	1
10. Phrenology	5	14	1
11. Physiognomy	11	3	6
12. Psycho-therapy	9	1	10
13. Absent Treatment	1	12	7
14. Gifted Children are "Peculiar"	12	8	0
15. Psychology Can Determine What Specific Job a Man Is Best Fitted for	16	4	0

needs devote some time and attention to evaluating the evidence for and against a large number of pseudo-psychological systems following this up with a presentation of what applied psychology can do and can not do. Thus it turns out that phychological instruction must serve as a sobering survey of the limitations of psychological applications. It may be that business men are free from such false beliefs yet the undersigned classifies himself as a "doubting Thomas."

University of Minnesota — DONALD G. PATTERSON.

BOOK REVIEWS

CLAPAREDE, EDOUARD. *Geneva, Switzerland,* Problems and *Methods of Vocational Guidance.* International Labor Office, 1922. Price, 40 cents.

Interest in vocational guidance has been growing so rapidly that the International Labor Office has taken cognizance of it as a subject of international importance. About two years ago the Office appointed the University of Geneva as its agent to collect information regarding the progress of the movement throughout the world, and to prepare news items for regular publication in the International Labor Review. This work has been carried on in increasing volume and with widening usefulness. In the meantime Professor Claparede has been conceiving and directing investigations fundamental to vocational guidance from his laboratory and from the Institute J. J. Rousseau, and he has brought out this brochure as a precipitation of the results of his study.

It is not a book. It is rather a survey that just escapes the severities of outline presentation. Furthermore it does not treat all phases of vocational guidance. Though it recognizes the importance of the social, economic, physiological and educational aspect it treats specifically only the psychological. This it envisages with great clearness. Particularly sane and proportionate is the critical discussion of the difficulties involved in the application of tests in practical vocational situations. In his summarized conclusions Professor Claparede gives what might be regarded as an international message:

"In conclusion it may be well to summarize very briefly the steps which should be taken to place vocational guidance on a firm foundation.

1. Psychological analysis of occupations and their classification according to the abilities required.

2. Determination of corresponding abilities in individuals and, as a preliminary, the testing of tests to determine their diagnostic, and even more their prognostic, value.

3. Investigation of the process of training and educability.

4. Determination of the importance of general intelligence in the various occupations.

5. It should never be forgotten that a forecast of ability will never be a certainty, but only a probability. Statistics should be prepared to determine this probability.

6. Efforts should be made to arrive at a uniform system of measurement so that psychographs or psychological diagrams may be comparable between different countries and between dif-

ferent vocational guidance offices. Grading by percentiles appears to be the most rational and least ambiguous method.

It is in the social interests of the community that every man should be in the position for which he is best fitted. It is also desirable and increasingly in accordance with the spirit of modern democracy that the various social functions should be allotted according to merit and not in virtue of any privilege. The only means of attaining this ideal is by a sound system of vocational guidance. All those who have at heart the harmonious organization of society, both employers' and workers' organizations, should therefore facilitate by all means in their power any investigation having for its object the establishment of industrial psychology upon a firmer foundation."

Every American phychologist who is touching ever so lightly, either in his teaching or in his investigation, upon vocational problems, should ponder on this brochure. He will find posed for him numerous problems for the experimental investigations so sorely needed, and he will find references to works of European investigators that are too little known in America.

Indiana University H. D. KITSON.

PRESCOTT, F. C. *The Poetic Mind.* New York: The Macmillan Company, 1922. Pp. 308, $2.00.

There are two classes of people to whom this book should be of special interest. The first are the psychologists. Professor Prescott has for the object of his study the working of the poet's mind. His evidence is from the most reliable source—namely, from the poets themselves. An English professor, he has collected from far and wide statements of the poets and other creative artists concerning the source and manner of their inspiration; and with a surprising unanimity of opinion he has found that they testify to a power within from which their creative works emanate and over which they have no voluntary control. This mysterious source of power Professor Prescott attempts to identify with the subconscious mind. He seeks to show further, always depending for his chief evidence on the poets themselves, that the same principles which apply to dreams according to Freud apply also to poems, which are emanations of the subconscious mind just as much as are nocturnal dreams. Poems, like dreams, have their origin in thwarted desires. Step by step Professor Prescott parallels the two kinds of dreams, showing that the impulse, the censorship, displacements, condensations, fusions, and fictional gratification are common to both and have the same explanations. Whatever be the merits of his findings, he has decidedly done the psychologists a service by bringing them into contact with a field of investigation which few of them are likely to be so well qualified as he *to explore—the mental operations during poetic inspiration* as explained by the poets themselves.

The book, however, should perhaps have even a greater appeal to the critic and the student of literature. In the first place, even if there is not a particle of truth in his thesis, it is a fascinating romance of the subterranean caverns of the human mind. In the second place, if his thesis is partly true, his work becomes valuable and suggestive. It sheds new light on such ancient themes as the relation of beauty to truth, the catharsis of tragedy, the universality of the greatest poetry, symbols and figures, and even puns. In the third place, it is very possible that his thesis is true. Once grant the existence of a subconscious mind of the kind accepted by many modern phychologists and the rest easily follows. The subconscious mind is old and wise, but not accessible to the average individual. The poet is one who for comparatively short intervals does have access to it, though through no volition of his own. What have been regarded as his vague and misty dreams are in reality flashes of truth from a deeper mind, which is racial, universal, unhampered by the limitations of time and space and immediate needs. Poetry, therefore, represents the highest wisdom of the race. Scientists plod along, and by the aid of the conscious mind finally confirm what the poet saw in his vision. Shelley, without the aid of modern psychology, states the idea exactly in his *Defense of Poetry*: "In the infancy of the world, neither poets themselves nor their auditors are fully aware of the excellency of poetry: for it acts in a divine and unapprehended manner, beyond and above consciousness, and it is reserved for future generations to contemplate and measure the mighty cause and effect in all the strength and splendour of their union."

It seems to me, however, that there is at least one rather serious inconsistency in the book. In Chaper VIII Professor Prescott is inclined to accept the conclusion of Wordsworth that emotions recollected in tranquility are "generally" the source of poetry. In Chapter XV he attempts to show that poetry is the result of thwarted desires: that, as the poet may not have the object of his desire, he seeks relief, even though it be a veiled relief, in a fictional gratification, as Keats, let us say, in *The Eve of Saint Agnes*. Such a relief is obviously needed and obviously would come when the emotion is powerful and not when it is recollected in tranquility. Likewise in the last chapter Professor Prescott has no basis for even suggesting the possibility of poets for seeing specific events in a literal sense, as Shelley foresaw his own death by drowning at sea.

Ohio University. B. L. JEFFERSON.

TERRY, PAUL WASHINGTON. *How Numerals Are Read.* Chicago, The University of Chicago, Supplementary Educational Monographs, No. 18, 1922. Pp. XIV+109. Price, $1.00.

Part I of Doctor Terry's excellent monograph deals with studies of the reading of numerals by introspective methods. Part II contains an account of the use of photographic apparatus in dealing with this problem.

It is the opinion of the reviewer that the chief value of the monograph lies in the suggested method of studying the problem rather than in the results of the experiments, since a very limited number of subjects participated. Part II, involving the use of photographic apparatus, contains the records of six subjects; in Part I, the number of subjects ranges in the various experiments from four to ten.

In the first experiment, seven arithmetic problems were solved by ten graduate students who kept notes on their experiences while reading the problems "with special reference to numerals." These observations were organized and recorded in a series of tables. The data made it possible to distinguish between a first reading and a re-reading in many instances as well as a partial and a whole first reading.

Other experiments in which the method of introspective analysis was employed investigated the range of correct recall of numerals after the first reading, undertook an analysis of the re-reading of numerals in arithmetical problems and studied the successes of four subjects in reading numerals in columns.

Important outcomes of the use of photographic apparatus are: "(a) The average number of digits included by a pause on numerals is decidedly smaller than the average number of letters included by a pause on words. (b) The average duration of pauses on numerals is greater than the average duration of pauses on words in the cases of all subjects. (c) The percentage of regressive pauses on numerals is greater than the percentage of such pauses on words."

The final chapter of the monograph contains very pertinent suggestions for teaching children to read arithmetic problems. Inasmuch as the reading of numerals requires closer attention on the part of the reader than does the reading of words, it is suggested that during the first reading pupils be taught to devote their entire attention to the conditions of the problem, developing, at the same time, skill in the partial reading of the numerals. These conclusions are supported by the findings of Gray and Freeman in the field of reading.

Ohio University. R. L. MORTON.

KNIGHT, FREDERICK B. *Qualities Related to Success in Teaching.* Teachers College, Columbia University Contributions to Education, No. 120, New York City. Pp. x+67.

When and why is a teacher a successful teacher? Many attempts have been made to answer this question in terms of such general categories as personality, poise, physical qualities, mental qualities, moral qualities, disciplinary skill, instructional skill, etc. The list can be extended and subdivided almost indefinitely from one of the several score cards for teachers now on the market. The objection that can be made to the above descriptive terms is their generality. Although they are more useful than a "general impression" as a guide to getting a useable criterion of teacher success, more exact and less subjective measures would be welcomed by the school administrator. The study here under discussion represents a contribution to scientific education in that it gives results and particularly an improved method of attack on this problem.

The first chapter summarizes the studies of Meriam, Elliott, and Boyce in this field. Chapters II, III, and IV are concerned with the data of the author's investigation and the relative significance of the various traits and qualities studied. The last chapter gives a trenchant criticism of Boyce's study.

The general procedure of the first part of the investigation was as follows. Ratings as to the relative teaching ability of 156 grade and high school teachers in three different school systems were obtained. That is to say, the best teacher was ranked first, the next best second, etc., when compared with all the other teachers of a given system. Three such rankings on the general ability to teach were secured from the supervisors, the teachers themselves, and a selected group of pupils respectively. By the method of correlation the validity of these rankings, and their relation to age, amount of experience, quality of handwriting, intelligence as measured by test, major academic interests, normal school scholarship, amount of professional study during active service, and ability to pass a professional test ("A Trade Test for Elementary School Teachers" by Knight and Franzen) was established.

A few of the more significant conclusions may be given.

"By using the coefficient of partial correlation we find, in the case of elementary school teachers, that, the factors of intelligence and normal school scholarship being constant, there is a mutual relationship of +.57 between ability to teach and ability to pass a professional test.... The amount of professional study accomplished during active service is also indicative of success in teaching.

"In the case of high school teachers, intellectual differences, as measured by mental tests, appear to be significant. For the selection of high school teachers the use of mental tests would be of value.

"These data, as a whole, may be interpreted to mean that the general factor of interest in one's work becomes the dominant factor in determining one's success in teaching....other measurable traits, either alone or in combinations, are not adequate explanations of teaching success.

"This study indicates that age, experience, quality of handwriting, intelligence as measured by tests, normal school standing, or the expressed interests of teachers are not closely related to success in teaching."

In the second part of the study intercorrelations were found among the traits teaching ability, ability to understand and manage people, intellectual ability, ability to manage things and mechanisms, general scholarship, skill in government or discipline, instructional skill, initiative, nervous and physical strength, ability to command the respect of pupils and general ability to get results, the teachers rating each other on these qualities after the connotations of these terms had been explained to them. Concerning these data the author concludes that the factor of general spread of estimate operates to invalidate such ratings. That is, "when a judge attempts to rate a teacher in some particular trait, his rating is a defense of his general estimate of that teacher, as well as a rating of the trait under consideration."

Doubtless the future will see a scientifically accurate and serviceable measure of teaching ability. The results and methods of such studies as this point the way to the solution of the problem.

Colorado College HERMANN REMMERS.

WALLIN, J. E. W., *The Achievement of Subnormal Children in Standardized Educational Tests.* Miami University Bulletin, Series XX, No. 7. $0.60.

In this bulletin of ninety-seven pages, Dr. Wallin gives the results of some educational tests administered by special class teachers to about 300 St. Louis subnormal children. The tests used were the Ayers spelling scale, lists A to F and I, L and O, and lists I and II in Starch's Spelling scale, Gray's oral reading and the spiral arithmetic excersises adapted from the Courtis tests.

Dr. Wallin states that this "study should have a two-fold significance for educators interested in the study and training of feeble-minded and backward children: first, it is the only survey thus far made by several standardized educational tests of pupils enrolled in special schools for mental defectives; second, most of the pupils tested had been psychologically examined and differentially diagnosed, so that it is possible to analyze the results from the educational tests with reference to the children's intelligence age, to their diagnostic classification, such as imbeciles, morons, borderline, backward, visual aphasia cases, etc., and to the grade assigned them by the special class teachers."

The results of the educational tests show that it is a waste of time and money to attempt academic instruction to low grade children below four years mentality and that little reading can be accomplished by a child under the mental age of six years. On the other hand the tests show that subnormal children who may develop to about the nine-year mental level should be taken care of in the public schools and that although their training should a reasonable amount of time should be devoted to academic be along the lines of the monumental arts and proper social habits, branches. Dr. Wallin is convinced that it is of extreme importance to take every precaution in placing children in schools and institutions for the mentally defective and that their educational attainment ability must be considered as well as a mental diagnosis made.

Ohio University. M. LA VINIA WARNER.

High Schools and Sex Education: A manual of suggestions on education as related to sex, Edited by Benjamin C. Gruenberg, Ph. D. Prepared under the direction of the Surgeon General, U. S. Public Health Service, in co-operation with the U. S. Bureau of Education. Washington, D. C., Government Printing Office, 1922. Pp. 98. Price $0.50.

This is a very helpful and practical book for teachers of all young people of secondary school age. It makes clear the need of better methods of training in matters of sex. In the very earliest pages it differentiates sex education from sex information—a point which the lay mind seems to have been slow in appreciating. Sex information without interpretation of its meaning to society and the individual and without inspiration to use it for the uplift of the race falls far short of its proper place in education. This little manual has been built up out of the accomulated experience of thousands of teachers who have attempted to handle the delicate subject of sex. It outlines the necessary preparation of teachers of sex education, discusses the type of personality necessary to make such teaching valuable and presents practical methods and devices by which sex education may be accomplished through the regular courses in biology, general science, physiology, physical education, home economics, and the social sciences. Numerous references are cited in order that the reader may have an adequate background for more detailed reading. In the brief appendices at the end of the book are some very valuable suggestions on the use of special lectures, exhibits, courses of study for teachers, personal hygiene and physical examinations. A very comprehensive bibliography of books dealing with sex and sex education ends the book. As a manual of what is being done and can be done in sex education this book deserves a place in every teacher's library.

Ohio University. G. H. GROWDON.

The World Book Company, Yonkers, N. Y., and 2123 Prairie Ave., Chicago, Ill., announces the following:

1. *Otis Self-Administering Tests of Mental Ability*, a new series of tests for grades from four to twelve and for college freshmen. The Higher Examination is for grades nine to twelve and college freshmen, the Intermediate for grades four to eight. Experimental try-outs have been given in preliminary editions. Results have given high correlations with teachers' marks and estimates. Four new advantages, it is argued, are to be had in these tests they are largely self-administering, very easily and quickly scored, afford a flexible time limit, and a wider variety of test material. There are two forms, A and B, each of equal difficulty. The Directions and Key with each package of 25, price $1.00 for Intermediate, $1.10 for Higher Examination.

2. *Pintner-Cunningham Primary Mental Test*: This is a new group test for the kindergarten, first and second grades and the only one on the market suitable for kindergarten pupils. It has been tried out in several previous editions. The advantage of a test applies to the kindergarten and primary grades are that it enables the teacher to classify beginners (with no previous school record), and also to discover by comparison of results, the degree of everlapping in mental ages.

The examination consists entirely of pictures. No reading or writing is necessary on the part of the pupils. The administering is sufficiently simple to require no special training on the part of the teacher.

The examination booklets are sold in packages of twenty-five at $1.45 net per package. The postpaid price of a Specimen Set is 20 cents. One Manual of Directions, one Class Record, one Report to Author and Percentile Graph are included in each package.

3. *National Intelligence Tests, Supplement No. 2.* The Committee on National Intelligence Tests of the National Research Council has prepared Supplement No. 2 to the Manual of Direc-

tions for the tests. This supplement contains revised age norms, based upon the examination of 32,972 pupils in nineteen communities; composition of material for the age norms; data showing variability of the average mental ability of children from city to city; mental age equivalent of scores; revised grade norms, with tables showing composition of the materials used, and a formula for finding IQ's and table for converting years and months into months.

Books and pamphlets for review should be sent to James P. Porter, Department of Psychology, Ohio University, Athens, Ohio.

BOBBITT, FRANKLIN. *Curriculum-Making in Los Angeles.* Supplementary Educational Monographs, No. 20, June, 1922. University of Chicago, Chicago, Ill. 106 pp.

BONNER, H. R. *Statistics of Teachers' Colleges and Normal Schools.* 1919-1920. Department of the Interior, Bulletin No. 8, 1922. Washington, D. C. 68 pp.

Consolidation of Schools in Maine and Connecticut. Department of the Interior, Rural School Leaflet No. 4, 1922. Washington, D. C. 12 pp.

COOK, KATHERINE M. *Supervision of Rural Schools.* Department of the Interior, Bulletin No. 10, 1922. Washington, D. C. 111 pp.

CRAFTS, WILBUR F. *That Boy and Girl of Yours—Sociology from Viewpoint of the Family.* The Baker & Taylor Company, 354 Fourth Avenue, New York. Price $1.75. 432 pp.

DEFFENBAUGH, W. S. *Administration of Schools in the Smaller Cities.* Department of the Interior. Bulletin No. 2, 1922. 75 pp.

DELGADO, HONORIO F. *El Dibujo de los Psicopatas.* Casilla 1589, Lima, Peru. 26 pp.

EDSON, NEWELL W. *Status of Sex Education in High Schools.* Department of the Interior, Bulletin No. 14, 1922. Washington, D. C. 12 pp.

FITT, A. B. *The Human Instincts in Business.* The Lothian Book Publishing Company Pty., Ltd. The Rialto, 497 Collins St., Melbourne, Australia.

FOLMER, ANTH AND GERRITSZ, J. *Tijdschrift Van Den Nederlandschen Werkloosheids-raad.* Bureau: Singel 530, Amsterdam.

GESELL, ARNOLD. *Exceptional Children and Public School Policy.* Yale University Press, New Haven, Conn. 66 pp.

GIVLER, ROBERT CHENAULT. *Psychology—The Science of Human Behavior.* Harper & Brothers, Publishers, New York. Price $3.00. 382 pp.

HEBB, BERTHA Y. *Value of the School Census.* Department of the Interior, City School Leaflet No 3. Washington, D. C.

HOKE, ROY EDWARD. *Improvement of Speed and Accuracy in Typewriting.* Johns Hopkins University Studies in Education, No. 7. Johns Hopkins Press, Baltimore, Md. Price 75 cents. 41 pp.

HOME ECONOMICS IN RURAL SCHOOLS. Department of the Interior. Home Economics Circular No. 13, May, 1922. Washington, D. C. 15 pp.

HOOD, WILLIAM R. *Review of Educational Legislation 1919 and 1920.* Department of the Interior, Bulletin No. 13, 1922. Washington, D. C. 30 pp.

JOHN, WALTON C. *Statistics of Agricultural and Mechanical Colleges for 1919 and 1920.* Department of the Interior, Bulletin No. 27, 1922. Washington, D. C. 29 pp.

Junior High School Mathematics. Course of Study and Syllabus. Cleveland Board of Education, Cleveland, Ohio. 45 pp.

KING, LE ROY ALBERT. *Status of the Rural Teacher in Pennsylvania.* Department of the Interior, Bulletin No. 34, 1921. Washington, D. C. 86 pp.

LATHROP, EDITH A. *How Laws Providing for Distribution of State School Funds Affect Consolidation.* Rural School Leaflet, No. 5. August, 1922, Department of the Interior. Washington, D. C. 12 pp.

LAUGHLIN, HARRY HAMILTON. *Eugenical Sterilization in the United States.* Psychopathic Laboratory of the Municipal Court of Chicago. Chicago, Ill. 502 pp.

LOMBARD, ELLEN. *Home Education by Means of Reading Course and the Co-operation of State and National Agencies.* Department of the Interior. Home Education Circular No. 2. Washington, D. C. 11 pp.

List of References on Rural Life and Culture. Library Leaflet No. 16, Department of the Interior, Washington, D. C. 10 pp.

PEAR, T. H. *Remembering and Forgetting.* E. P. Dutton & Company, 685 Fifth Avenue, New York. Price $3.00. 242 pp.

Professional Education of Teachers in Cleveland. Cleveland Board of Education, Cleveland, Ohio. 92 pp.

Record of Current Education Publications. Department of the Interior, Bulletin No. 33, 1922. Washington, D. C. 30 pp.

Reorganization of Home Economics in Secondary Schools. Department of the Interior, Bulletin No. 5, 1922. Washington, D. C. 38 pp.

Statistics of Universities, Colleges and Professional Schools 1919-20. Department of the Interior, Bulletin No. 28, 1922. Washington, D. C. 147 pp.

Surveys of the Department of Instruction, Cleveland Public Schools. Cleveland Board of Education, Cleveland, Ohio. 29 pp.

THWING, CHARLES FRANKLIN. *Higher Education in Australia and New Zealand.* Department of the Interior, Bulletin No. 25, 1922. Washington, D. C. 44 pp.

VALDEZAN, HERMILIO AND DELGADO, HONORIO F. *Revista de Psiquiatria y Disciplinas Conexas.* Publication Primestral Vol. IV No. 3. Casilla 1589, Lima, Peru. 262 pp.

WILIAMS, J. HAROLD. *The Twenty-Four Hour School.* Educational Research Bulletin, Vol. I, No. 6, Pasadena City Schools, Pasadena, Calif. 8 pp.

WINDES, EUSTACE E. *A Plan for the Organization of a County System of Agricultural Instruction in Elementary Rural Schools.* Department of the Interior. Rural School Leaflet No. 6. Washington, D. C. 8 pp.

ZOOK, GEORGE F. *National Conference of Junior Colleges, 1920.* Department of the Interior, Bulletin No. 19, 1922. Washington, D. C. 73 pp.

ZOOK, GEORGE F. *Report on the Higher Educational Institutions of Arkansas.* Department of the Interior, Bulletin No. 7, 1922. Washington, D. C. 18 pp.

THE
JOURNAL OF APPLIED PSYCHOLOGY
VOL. VII JUNE, 1923 No. 2

FACIAL EXPRESSION IN ADVERTISEMENTS

By Harold E. Burtt and J. Camden Clark
Ohio State University

INTRODUCTION

The facial expression portrayed in an advertising cut is frequently an important factor in the sale. Efforts to portray "satisfaction" are quite common, and it is to this feeling that the present study is confined.

Experiments upon the effectiveness of various types of advertising appeals have shown that this effectiveness varies with different commodities. This suggests that a similar relation may hold in the case of facial expression and that perhaps a satisfied expression in a cut may have quite different value in advertising different varieties of goods. It would seem, a priori, that certain purchases are serious and that pronounced satisfaction in the facial expression would be misplaced, while with other objects whose use normally provokes considerable satisfaction, such an expression would be desirable. It is probable that the advertiser sometimes errs in the use of such facial expressions. One notes, for instance, cuts of a man laughing moderately at a game of solitaire or smoking a choice cigar with marked solemnity. Perhaps this is correct or perhaps it is irrelevant but at least it suggests a problem.

The following study is devoted mainly to the problem of whether the optimal degree of satisfaction portrayed in the facial expression of an advertisement varies with the commodity advertised. A subsidiary problem is whether the apparent satisfaction in the facial expression of a photograph corresponds to the actual satisfaction felt by the subject who posed for it, i. e., whether the picture actually conveys the feelings intended.

It is impossible to approach such problems by strictly objective methods. The only possibility is to obtain observers' judgments as to how they would react to certain pictures with reference to the purchase of various commodities. Judgments of this sort are perhaps difficult and none too reliable but, with a sufficient number of observers and proper control of material and conditions, are undoubtedly more valuable than a priori generalization.

MATERIAL.

Two sets of material were used in the experiment. The first consisted of a series of bust photographs of the same male subject showing different facial expressions. Considerable preliminary work was necessary in securing a subject who could pose properly. The set finally used consisted of six photographs that were taken in succession, with the subject in the same position. The lighting, exposure and other technical aspects were constant. The subject posed in succession six increasing degrees of satisfaction starting with slight and ending with rather extreme. He was confident of the climactic effect and corroborated this by ranking the pictures subsequently.

A set of these photographs was enlarged to 7x9 inches and each one mounted without margin under celluloid for daylight display. Contact prints 4x5 inches were mounted individually on white cardboard of suitable size for projection in a Balopticon.

The other material consisted of 23 cuts of men's faces taken from current advertisements and portraying various degrees of apparent satisfaction. It was thought best to limit the faces to those of one sex thus avoiding an additional variable. The advertisements were trimmed so that none of the copy remained, and so that nothing in the cut suggested the kind of commodity advertised. It was impossible to control size but all the cuts were sufficiently large and sufficiently well printed to be seen without difficulty when projected on the screen. Preliminary experiments served as a basis for selecting this final set of cuts from a somewhat larger number. The cuts were centered on white cardboard of suitable size for projection.

For presenting the enlarged photographs a display board was used. This consisted merely of a medium gray cardboard about 2x3 feet mounted on a light framework with the longest dimension horizontal. There was a narrow gray rack at the bottom of the card and the top was tilted slightly away from the observer so that the photograph could rest against the background. For the small photographs and the cuts a Balopticon was used to project them upon a screen. A metal plate actuated by a spring held the card on which the picture was mounted in position and to change material it was necessary merely to depress the plate and insert another card.

METHOD.

Series I employed the enlarged photographs and the display board. The experiments were made in daylight with one observer at a time and all the results recorded by the experimenter. This series comprised three parts, which were given in succession.

In *Series IA* the photographs were presented one at a time in random order (different for each subject) with the instructions, "Describe fully the feelings which this picture suggests to you." The complete introspection was recorded and subsequently analyzed.

In *Series IB* the observer was given the six photographs in a pile arranged in random order and was told: "These pictures were taken when the subject felt different degrees of satisfaction. Arrange them in order with the one showing the most satisfaction on this end and the one showing the least satisfaction on this end." The observer's ranking was recorded in terms of the key letters on the backs of the photographs.

In *Series IC* the photographs were again shown singly in random order with the instructions: "Name five commodities for which you think this picture would be effective in an advertisement. For which of these five would it be most effective, next most effective, etc.?" The five commodities mentioned were recorded in order for each photograph.

Series II practically repeated series I using the group method of experimentation. The treatment of results was identical. The smaller photographs were projected on the screen. The room was, of course, dark but could be illuminated sufficiently for writing by merely opening a door in the side of the Balopticon. The pictures were each given a key letter and presented in the same order in all three parts of the series. This order began and ended with modrate degrees of satisfaction and no two adjacent degrees occurred in succession.

In *Series IIA* and *Series IIC* the instructions were identical with those of *IA* and *IC*. A picture was presented and its code letter given (verbally). After thirty seconds of darkness the room was illuminated for thirty seconds during which time the observers recorded their introspection on a blank. Another similar alternation of darkness and illumination followed for the benefit of those who had not finished or who wished to note further aspects. A third alternation was given if requested by any observer.

Series IIB involved ranking the photographs. Previous to the experiment the following six degrees of satisfaction had been written on the blackboard, but the observers' attention not called to them until the proper time.

1. Extremely satisfying.
2. Very satisfying.
3. Quite satisfying.
4. Fairly satisfying.
5. Slightly satisfying.
6. Very slightly satisfying.

The instructions were as follows: "These pictures were taken when the subject felt different degrees of satisfaction. These degrees are as follows, (experimenter points to the list on board). You are to write down the six letters and the six numbers with the proper letter. I will show the pictures in succession several times so that you can decide carefully." The photographs were then shown repeatedly in the same order and the code letter stated for each as it was shown. Each presentation was of fifteen seconds duration followed by a few seconds illumination.

Series III employed the six small photographs and the twenty-three cuts in a group experiment. In the discussion of results the former will be designated *IIIA* and the latter *IIIB*. They were given, however, at one sitting with the photographs in a random order followed by the cuts in a random order. These series were concerned only with the problem of commodities and the judgments were controlled rather than spontaneous. It had been found that practically all the commodities mentioned spontaneously in *series IC* and *IIC* could be summarized under ten classes (cf. infra, table II). Accordingly, a mimeographed blank was provided with ruled columns headed with the names of these classes. Successive rows were likewise ruled and numbered at the left from 1 to 29. The blanks were distributed and the following instructions given. "This is an experiment upon facial expressions in relation to advertising. I am going to show you a series of pictures and in each case you are to judge for what classes of commodities the picture would be effective in an advertisement. By effective is meant whether it would induce persons to buy the commodity in question. Confine your answers to the classes listed on the blank. After the number of the picture put a 1 in the column to indicate the commodity for which it would be most effective and 2 in the column to indicate second choice. You may mark other choices if you wish but always give at least two."

A picture was presented and its number given. After fifteen seconds the room was illuminated for fifteen seconds, (an assistant operating the wall switch) for the observers to record their judgments. Another alternation of picture and illumination followed. The experimenter then queried, "All right?" and if there was no request for further repetition proceeded to the next picture. There was no break between *series IIIA* and *IIIB*.

The ranking of satisfaction portrayed in the photographs was obtained from the subject who posed. As this was impossible with the cuts, fifteen observers who did not participate in *series III*, ranked them. These observers were individually given the twenty-three cards and told to arrange them with the one showing the most satisfaction on the top of the pile and the one showing the least on the bottom and the intervening ones in order. The pile was returned to the experimenter who recorded the order. Then the fifteen ranks given each picture were averaged. There was some disagreement among the judges in their rankings of the various cuts. The mean variation of the fifteen ranks assigned to each picture was computed. The smallest of these mean variations is 1.6, the largest 6.7, and the average 4.2. If two pictures in this series had each this mean variation of 4.2, the probable error of difference would be 1.3. So it would seem that a given picture was located at least within a few places of its correct position.

The subjects were students—mostly undergraduates—in psychology classes at Ohio State University. *Series I* had 9 subjects, *series II*, 16, and *series III*, 67. The group experiments were conducted by the senior writer, (Burtt).[1]

RESULTS.

The minor problem above mentioned will be discussed first, viz., the correspondence of satisfaction actually felt with that apparent in the facial expression. *Series IB* and *IIB* give the most direct answer to this problem. For convenience in subsequent discussion the six photographs will be denoted by the letters U, V, W, X, Y, and Z, with U the one in which the subject who posed felt the least satisfaction, V the next degree and so on up to Z the greatest satisfaction. The extent to which the observers' rankings of the six photographs correspond to this order of satisfaction felt by the subject is shown in Table I.

TABLE I.

AVERAGE RANKS.

Series	U	V	W	X	Y	Z
IB	5.8	4.7	4.4	2.6	2.2	1.2
IIB	5.9	4.9	4.0	2.4	2.4	1.3
Both	5.8	4.8	4.2	2.5	2.3	1.3

The original ranks were tabluated with the six made by a given observer in a row. Then the columns were averaged to obtain the average rank assigned each photograph. These averages appear in table I, for *series IB*, *IIB* and both together.

It is obvious that the observers judged quite accurately the satisfaction felt by the subject of the photographs. Considering the averages of both series the correlation is perfect. Moreover, the more detailed results show that all but three of the twenty-five subjects ranked U lowest and all but two ranked Z highest. The spontaneous descriptions of feelings suggested by the photographs indicate a rather climactic order from U to Z.

This suggests Langfeld's[2] finding,—covering a wide range of emotions and attitudes but not the one specifically studied in the present case,—of "uniformly good and consistent" judgments of emotions from facial expression.

Journal of Abnorm. Psych. 1918, 13, 172.

---

[1] The writers are indebted for some suggestions during the earlier stages of the work to Dr. Harry W. Crane.

[2] Langfeld, H. S.: *The Judgment of Emotions from Facial Expression.*

The main problem of the experiment was the effectiveness of various degrees of apparent satisfaction for advertising various commodities. The results will be considered in terms of the 10 classes of commodities under which practically all the spontaneous judgments fell. The results for *series IC* and *IIC* were combined, as both used spontaneous reports on the same six photographs. A preliminary tabulation was made to show the number of times each commodity was mentioned in connection with each picture. Separate tables were made for first choices, second choices, etc., and total choices. *Series IIIA* which involved the same material with controlled judgments was tabulated in the same way. In *series IIIB* it did not seem worth while to evaluate each picture separately especially in view of the rather considerable disagreement among the fifteen observers whose judgments provided the original rankings for satisfaction. To facilitate comparison with the other series of six photographs the cuts were arbitrarily combined into six groups on the basis of the average satisfaction rankings. Each group comprised 4 cuts with the exception of the lowest satisfaction group which comprised 3 cuts. The division was made in this way to avoid drawing the line between cuts with the same average rank. A preliminary tabulation was made to show the number of times each commodity was mentioned in connection with each of these six groups of cuts. Separate tables were made for first choices, second choices and total choices.

The results could now be considered from two standpoints. On the one hand it was possible to determine for each commodity what proportion of the mentions were accorded to each picture or group of pictures, i. e., what degree of satisfaction was most frequently selected as appropriate for a given commodity. On the other hand, it was possible to note for each picture or group of pictures what proportion of the mentions were accorded to each commodity, i. e., what commodities were most frequently selected as ones for which a given degree of satisfaction would be appropriate.

Table II presents the results from the former of the above standpoints.

TABLE II.

PERCENTAGE DISTRIBUTION OF TOTAL CHOICES FOR EACH COMMODITY.

Commodity	Series	n	U	V	W	X	Y	Z
Amusements—indoor	IC+IIC	33	3	3	12	21	27	33
	III A	87	1	25	22	23	14	15
	III B	287	3	12	31	16	22	17
	Av.		2	13	22	20	21	22
Amusements—outdoor	IC+IIC	14	0	14	7	29	21	29
	III A	98	0	13	12	32	26	17
	III B	313	1	11	19	30	27	12
	Av.		0	13	13	30	25	19
Automobiles	IC+IIC	15	20	27	20	20	7	7
	III A	55	4	13	15	20	33	16
	III B	276	8	7	29	23	7	26
	Av.		11	16	21	21	16	16
Clothing	IC+IIC	94	24	14	18	15	18	11
	III A	106	51	18	13	4	5	9
	III B	425	32	16	18	21	3	10
	Av.		36	16	16	13	9	10
Food	IC+IIC	37	0	0	16	27	32	24
	III A	70	0	11	21	16	26	26
	III B	253	3	23	14	12	29	18
	Av.		1	11	17	18	29	23
Photographs	IC+IIC	18	44	11	22	22	0	0
	III A	127	51	15	9	7	9	8
	III B	205	31	22	11	11	10	14
	Av.		42	16	14	13	6	7
Publications	IC+IIC	19	21	0	5	26	11	37
	III A	33	3	9	30	12	12	33
	III B	254	15	28	13	6	13	25
	Av.		13	12	16	15	12	32
Schools	IC+IIC	22	55	14	18	0	9	5
	III A	23	26	4	26	4	22	17
	III B	230	13	46	7	10	13	10
	Av.		31	21	17	5	15	11
Smoking supplies	IC+IIC	48	4	12	17	25	19	23
	III A	84	4	30	15	21	17	13
	III B	312	6	12	9	16	15	42
	Av.		5	18	14	21	17	26
Toilet articles	IC+IIC	92	4	7	11	33	20	26
	III A	67	2	13	10	25	18	31
	III B	352	15	10	23	14	29	9
	Av.		7	10	15	24	22	22

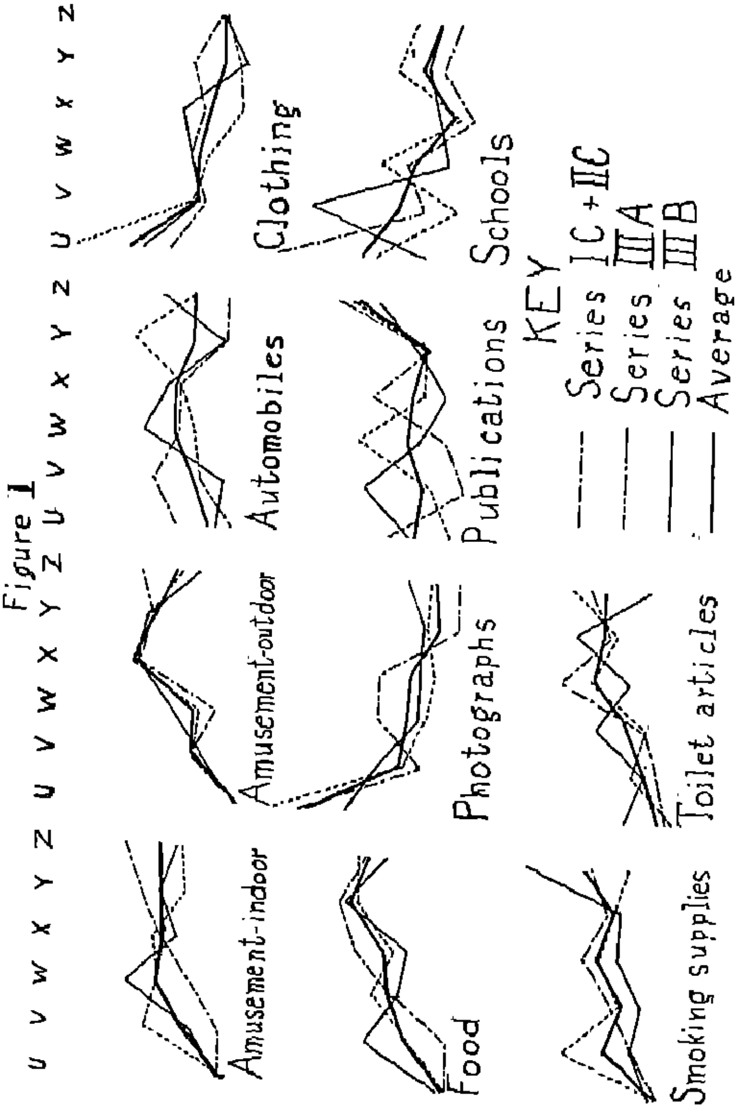


Figure I

U V W X Y Z
Amusement-indoor
Amusement-outdoor
Automobiles
Clothing
Food
Photographs
Publications
Schools
Smoking supplies
Toilet articles
KEY
Series IC + IIC
Series IIIA
Series IIIB
Average

The original figures were reduced to per cents to make the various series comparable. Only "total choices" are considered. First choices were tabulated separately but did not differ substantially from total choices and the latter were then selected because they involved a larger number of cases. The first column gives the classes of commodities, the second the series and the third the number of choices or mentions involved. The following six figures in a given row indicate the per cent of mentions for each of the six pictures or group of pictures. The headings U, V, W, etc., indicate increasing degrees of apparent satisfaction from U to Z; for *series IC, IIC* and *IIIA* they denote the six photographs, and for *IIIB* they denote groups of four cuts except that U involves only three cuts. For example, of the thirty-three times indoor amusements were mentioned in *series IC* and *IIC*, three per cent of these mentions were for photograph U, three per cent for V, twelve per cent for W, etc; while of the 287 mentions of indoor amusements in *series IIIB*, three per cent were for U, twelve per cent for V, thirty-one per cent for W, etc. Moreover, the average of the three series as far as indoor amusements are concerned gives two per cent for U, thirteen per cent for V, etc.

These same results are shown graphically in Figure 1.

The ten sets of curves correspond to the ten blocks in Table II. In each set the different lines denote the three series and the average of the three as shown in the key. The abscissae of the curves denote from left to right the six classes U, V, W, X, Y, Z, while the ordinates represent per cents. The total of the ordinates of each curve is, of course, 100 per cent. It should be noted in the figure and the table that the value for U in *series IIIB* should be perhaps a trifle larger as it involves three pictures instead of four.

The results are considered from the other aforementioned standpoint in Table III.

TABLE III.

PERCENTAGE DISTRIBUTION OF TOTAL CHOICES FOR EACH PICTURE.

Picture	Series	Amusements indoor	Amusements outdoor	Automobiles	Clothing	Food	Photographs	Publications	Schools	Smoking supplies	Toilet articles
U	IC+IIC	2	0	5	40	0	14	7	21	4	7
	III A	1	0	1	41	0	49	1	4	2	1
	III B	2	1	6	35	2	17	10	8	5	14
	Av.	2	0	4	39	1	27	6	11	4	7
V	IC+IIC	3	5	11	35	0	5	0	8	16	16
	III A	17	10	6	15	6	15	2	1	20	7
	III B	7	7	4	14	11	9	14	21	7	7
	Av.	9	7	7	21	6	10	5	10	14	10
W	IC+IIC	7	2	5	29	10	7	2	7	14	17
	III A	16	10	7	12	13	10	9	5	11	6
	III B	17	11	15	15	7	4	6	3	5	16
	Av.	13	8	9	19	10	7	6	5	10	13
X	IC+IIC	8	5	3	16	11	5	6	0	13	34
	III A	16	25	9	3	0	7	3	1	14	13
	III B	10	19	13	18	6	5	3	5	11	10
	Av.	11	16	8	12	9	6	4	2	13	19
Y	IC+IIC	12	4	1	23	16	0	3	3	12	25
	III A	10	20	14	4	14	10	3	4	11	10
	III B	13	17	4	3	15	4	7	6	10	21
	Av.	12	14	6	10	15	5	4	4	11	19
Z	IC+IIC	14	5	1	13	12	0	9	1	14	31
	III A	10	14	7	8	15	8	9	3	9	17
	III B	9	7	14	8	9	5	12	4	25	6
	Av.	11	9	7	10	12	4	10	3	16	18

The first column gives the picture or group of pictures and the next the series. The following columns in any row indicate the per cent of mentions for each of the ten commodities. For instance, with picture U in *series IC* and *IIC*, 2 per cent of the observers selected indoor amusements as a commodity for which U would be suitable, 5 per cent selected automobiles, 40 per cent clothing, etc.

Series IC and *IIC* involved about fifty total mentions, *series IIIA* about 130 and *IIIB* about 500, for each photograph or cut. Figure 1 and Table III show fairly consistent results with certain commodities but not with others.

The curves for indoor amusements agree in starting low and they approach each other closely again at X. In Table III this commodity is seldom mentioned for U but more frequently thereafter. Apparently slight satisfaction is undesirable while it is not clear whether moderate or greater degrees are most desirable.

With outdoor amusements the case is clear. The curves show marked agreement with a minimum at U and a maximum at X. In the table for X, outdoor amusements stand next to the highest in average mentions. Evidently a rather marked but not extreme degree of satisfaction is the most desirable.

It is impossible to draw any conclusion regarding automobiles because there is little agreement among the curves and the average curve is nearly level.

The curves for clothing all start high and descend first abruptly and then gradually, while in the table, picture U has a large proportion of mentions as suitable for clothing. Evidently, the slightest degrees of satisfaction used in the experiment were the most suitable for this commodity.

Food shows a fair agreement in the curves, with a gradual rise toward the higher degrees of satisfaction. This relation is not as clearly manifest in the table, but on the whole it seems that a fairly high degree of satisfaction should be effective in food advertisements.

The curves for photographs are rather similar to those for clothing and the tabular results also are similar. Apparently, the slight degrees of satisfaction should be most effective.

The curves for publications are equivocal with a suggestion of a uniform rise at the end. This fact is not clearly borne out by the table.

There is considerable discrepancy at the left in the curves for schools although the average trend is downward. The table is inconclusive. If there is any suggestion it is toward the effectiveness of the lesser degrees of satisfaction.

Smoking supplies show a fairly consistent and gradual rise in the curves. The table indicates frequent mention of this commodity, for all the pictures except U. It would seem that any but the slight degrees of satisfaction would be desirable and perhaps there would be a slight gain in effectiveness as the satisfaction was increased.

Toilet articles have a fairly consistent rise through the lesser degrees of satisfaction with little average change thereafter. In the table toilet articles are frequently mentioned for all pictures except the first. It seems that an average or greater degree of satisfaction should be effective in advertising such articles.

The above results depend of course upon the judgments of an untrained, unselected group of students. They may have judged in some instances by factors other than the one studied. Recognition of a picture, for instance, might sometimes affect the results. the conclusion that for clothing and photographs, advertisements showing rather slight amounts of satisfaction should be more effective than those showing greater amounts, while for outdoor amusements, food, smoking supplies and toilet articles the moderate or high degrees of satisfaction should prove more effective. In the case of outdoor amusements and food there is a further suggestion that the satisfaction, if carried to the extreme, may lose somewhat in effectiveness.

A further interesting problem would be to determine whether current advertising seemed to follow the above principles or to run counter to any of them. This would necessitate the collection of a large number of advertising cuts and their ranking by a group of observers. A little evidence may be obtained by considering the twenty-three cuts used in the present experiment in the light of the commodities they had been originally used to advertise. Table IV shows how the commodities originally advertised by the cuts used are distributed through the six degrees or groups of satisfaction. There were, of course, too few cuts for a given commodity to permit much generalization, but clothing, for instance, occurs only in classes U and V which was shown above to be a desirable situation. Outdoor amusements avoid U and V as they ought to. Food, smoking supplies and toilet articles for the most

TABLE IV.

	U	V	W	X	Y	Z
Indoor amusements					1	1
Outdoor amusements			1	1		
Clothing	2	1				
Food		2	1	3	1	2
Smokers' supplies		1				
Toilet articles	1		2		2	

part occur in the higher degrees but have some instances in U and V which according to the above findings is undesirable. Without more data on this point it would be rash to criticize advertising policies but the method just described if applied on a large scale would doubtless be illuminating.

The results of the main portion of the experiment are, of course, applicable only to the extent that the facial expression of a cut is a selling point in itself, i. e., the apparent satisfaction of the user is supposed to be an inducement to the reader to buy. There are many occasions when there is some other leading selling point and the facial expression is one designed to emphasize it. In such a case, the foregoing results would be irrelevant. At any rate it seems clear that the advertiser should not always strive to portray a maximum satisfaction in facial expressions but that the optimal degree varies with different commodities.

SUMMARY.

The foregoing experiment employed photographs posed by a subject feeling different degrees of satisfaction, and cuts from current advertisements portraying apparently various degrees of satisfaction. Individual and group methods were used. The observers stated the feelings suggested by the pictures, ranked them for apparent satisfaction and gave the commodities for which they thought each picture would be most effective in an advertisement. The results were tabulated to show the correspondence between actual satisfaction felt by the subject posing for the photographs and apparent satisfaction judged by the observer. The correlation proved to be very high.

The results were also treated in various ways to determine the optimal degree of apparent satisfaction for advertising different commodities. As far as the results go, the lesser degrees of satisfaction appear more suitable for advertising clothing and photographs, while the moderate or high degrees are better for amusements, food, smokers supplies and toilet articles.

The results are, of course, irrelevant when the facial expression of satisfaction is not itself a selling point, but in cases where it is a selling point the conclusion may be drawn that it is undesirable to always strive for the utmost satisfaction but that the optimal amount depends on the kind of commodity advertised.

COMPARATIVE SOCIAL TRAITS OF VARIOUS RACES, SECOND STUDY

By C. B. Davenport and Laura C. Craytor.

In "School and Society" for October 22, 1921, one of us published a study of "Comparative Social Traits of Various Races." This study was based on only 51 individuals and was confessedly fragmentary. It was published in the hope that it would stimulate others to make similar studies. How far this hope has been realized we can not say, but at least the first paper doubtless made it easier to secure cooperation of other teachers in collecting such data. The present study is due to the interest taken by Dr. R. S. Benedict of the Stuyvesant High School, who undertook to secure additional data from his pupils and to whom we are greatly indebted. These have doubled our totals. Since the latter are exclusively for boys, while the earlier were exclusively for girls it has seemed useful to combine the two collections to get results that may be expected to hold for the population as a whole. A comparison of the sexes is hardly possible with the small numbers available.

As stated in the earlier paper the purpose of the study is to replace opinion as to race differences in social traits by quantitative estimates.

The method is exactly the same in the two studies. The independent judgments of one to three teachers were obtained concerning ten traits of each pupil, using the graphic rating scale of the Scott Company, Philadelphia ("School and Society" XIV, p. 345). The data were tabulated and appraised as stated in the earlier paper. Probable errors have been calculated in order that the significance of the differences may be made clearer.[1] The results are tentative. It is hoped that this paper may stimulate to further studies of this subject.

RESULTS.

Altogether 188 sets of judgments were rendered on 102 persons, as follows: 20 on 10 Germans; 15 on 9 Irish; 37 on 22 Italians; 59 on 31 Austrians, and 57 on 30 Russians, also 116 on 61 Jews, being the sum of the Austrians and Russians, just listed.

There were calculated for each of these nationalities the mean of the judgments for each trait and the probable error of the mean. These means and probable errors are given in Table I. Thus under Germany, Leadership, we find the mean of the 20 estimates made on 10 persons to be 59.7±3.0. This means that the position of students of German stock, in respect to leadership, was on the average slightly above mediocrity, or 50%. There is, however, an even chance that the true value (theoretically based on an infinite number) is only equally apt to occur within the range

56.7—62.7 as outside that range. It is usually held that differences between two means which are greater than three times the square root of the sum of the probable errors of the means is certainly a significant difference; that it depends not merely upon *small numbers*, but upon a real difference in the "set" of these numbers. For example, while the mean of 20 estimates concerning leadership of students from Germany is 59.7±3.0, that of 15 estimates of students born in Ireland is 38.3±5.0. There is, thus, a difference of 21.5 in the mean of the judgments for the two countries. The square root of the sum of the squares of the probable errors is 5.8. Three times 5.8 is 17.4 which is less than the difference of the means; hence the *difference* between the means is probably a biologically significant one. Even with the probable errors the "conclusions" must be held to be merely tentative; for the "probable error" has not the significance with small numbers that it has with large numbers.

TABLE I.

MEANS AND THEIR PROBABLE ERRORS.

Trait.	No. I Germany.	No. 2 Ireland.	No. 3 Italy.	No. 4 Austria.	No. 5 Russia.	No. 6 Jews.	Average of Means (unweighted)
I. Leadership.	59.7±3.0	38.3±5.0	37.7±3.2	46.6±2.6	44.7±2.4	45.6±1.8	45.4
II. Pertinacity.	67.2±3.2	55.0±4.3	52.8±2.8	58.8±2.0	57.6±2.0	58.2±1.4	58.3
III. Humor.	62.0±3.4	55.8±4.8	49.1±3.0	54.9±2.3	49.6±2.4	52.3±1.7	54.3
IV. Frankness.	56.5±2.8	48.6±3.8	47.7±2.6	55.1±2.0	50.9±2.0	53.0±1.4	51.8
V. Suspiciousness	38.4±3.3	46.6±2.2	42.4±2.0	45.8±1.8	47.7±1.5	46.7±1.2	44.2
VI. Sympathy.	54.7±1.8	42.3±4.4	48.2±2.0	48.6±1.8	51.3±1.5	49.9±1.2	49.0
VII. Loyalty.	79.8±2.4	76.0±2.2	77.9±1.2	71.1±1.5	74.6±1.3	72.8±1.0	75.9
VIII. Generosity.	75.8±2.2	72.7±1.8	73.7±2.1	71.4±1.6	72.7±1.9	72.0±1.2	73.3
IX. Obtrusiveness.	47.3±2.6	44.0±3.2	41.0±2.5	51.3±2.0	52.7±1.9	51.9±1.4	47.3
X. Coolness.	67.0±2.3	66.3±3.6	58.4±2.5	56.7±2.0	56.3±2.1	56.5±1.4	60.9
No. of Estimates.	20:10	15:9	37:22	59:31	57:30	116:61	

In Table I the means and probable errors are given for five nationalities, namely, pupils who came, or whose ancestors came, from Germany, Ireland, Italy, Austria and Russia. The students of Austrian and Russian stock in this New York City school are almost exclusively Jews. It seemed worth while to combine those from Austria and Russia under the heading of "Jews", and these results are given in column 6. Table I then should give us some sort of an answer, though imperfect, to the inquiry with which we started as to the difference in social traits of persons of different nationalities.

CONCLUSIONS.

While numbers are still insufficient to form final conclusions, the results suggest the following tentative conclusions.

In *leadership* the German children stand first at 59.7±3.0 point quite markedly above the others. The difference between this mea and that of the Jews is less than three times the probable erro Nevertheless the Jews may be considered as a race characterize by a medium leadership. On the other hand, the Irish and Italian are characterized by a low standard in leadership, and the difference between them and the middle group is, doubtless, significant.

In *pertinacity* the Germans stand first. This group is probably higher in respect to pertinacity than the Jews. The Irish and Italians stand lowest in this respect.

In *humor*, the German children stand first (62) and the Irish stand second. However, the difference between these two nationalities is not great and, indeed, not to be differentiated from the Austrians and the Jews in general. Between this group and the Italians there is probably a real difference in respect to national humor.

In *frankness* the German children again stand first (57) and the Jews, and especially the Austrians, are a close second. The Irish and Italians fell into a slightly lower class.

In *suspiciousness*, Jews and Irish stand above the others. Probably the Italians are not significantly lower. The Germans, however, are probably least in this respect and certainly are quite distinct from the Jews.

In *sympathy* the Germans stand highest, the Jews and Italians may be included in the medium category. The Irish are least in this respect.

In *loyalty*, the Germans stand highest, but in the same category are also the Italians and Irish. The Jews occupy a lower position.

In *generosity* the Germans, Italians, Irish and Jews stand at nearly the same level.

In *obtrusiveness* the 61 Jews stand at the top (52) and seem to form a class distinct from the Germans, Irish and Italians which form a middle or lower group.

In *coolness* the Germans and Irish stand easily first (86). In a significantly lower group stand the Italians and Jews.

The 10 Germans stand at, or near the top in Leadership, Pertinacity, Humor, Frankness, Sympathy, and Loyalty, more than half the traits. They occupy a relatively low position in Suspiciousness.

The 9 Irish fall in the upper group in respect to Humor, Suspiciousness, Loyalty and Generosity, and in the lowest group in respect to Leadership, Pertinacity, Frankness and Sympathy.

The 22 Italians stand in the upper group in regard to Loyalty and Generosity and lowest in regard to Leadership and Humor.

The 61 Jewish children stand in the first group in Humor, Suspiciousness, Generosity, and Obtrusiveness and in the medium group for the other traits.

In addition to the mean values of the social traits for the different nationalities, there is a difference in the variability of judgments of persons from different countries, and it has seemed worth while to consider this variation. Accordingly, the standard deviations of the judgments for each nationality and trait are given in Table III. The probable errors have also been calculated as a means of judgment of significance, or rather, lack of significance of the difference of these indices of variability. Thus, from Table III, it appears that the variability in the judgments on Leadership is greatest for the Jews the standard deviation being 28.1±1.3. It is least for the Germans, 20.0±2.1. Although the difference may not be a significant one, there is a suggestion that the Jews (as represented in our statistics) include strains that vary more from each other in Leadership than in the case of the Germans.

On the other hand in Pertinacity, the Italians showed themselves the most variable and the Irish the least, although the difference, even in this case, is less than three times the probable error. Again in Humor the variability in judgments rendered on Italians, Jews and Irish are high, and for Germans somewhat lower. One of the striking differences in Table III is the slight variability of the Germans in regard to Coolness (15.2), a trait in which they stand at the top, and the variability of the Jews in this trait (22.8), a trait in which they stand at the bottom. Taking the average of the standard deviations of the ten traits for each of the nationalities, we find that the variation is greatest in the Jews, (21.8) and next in order in the Italians and Irish and least in the Germans.

From another standpoint it is interesting to compare the variability of the different traits for all nationalities. This average has been entered in the right hand column of Table III. The order of the average variation for these ten traits is as follows: Leadership—26.5, Humor—25.6, Pertinacity—23.2, Frankness—21.6, Coolness—20.7, Obtrusiveness—20.4, Sympathy—18.2, Suspiciousness—17.7, Generosity—16.2, Loyalty—14.2.

The interpretation of the differences in averages of the ten traits is not easy. It may indicate a difference in the intrinsic variability of the traits or a difference in the ease or certainty of estimating them. One sees no inherent reason why it should be easier for teachers to estimate Loyalty of pupils than to estimate capacity for Leadership. It seems to me more probable that Loyalty is a trait in which students (in their teachers' opinions) stand uniformly rather high, and that Leadership is a quality in which they show great diversity.

---

[1] Dr. J. Arthur Harris calls my attention to the fact that the probable errors and the standard deviations upon which they are based are probably influenced, to some extent, by the fact that different teachers rated the different individuals; and one set of teachers rated a larger proportion of the German students than another set which treated, say, a larger proportion of the Irish. Thus any difference in the personal equation of these two groups of raters would be reflected in the magnitude of the variabilities. Representatives of the different races were, of course, distributed among the teachers without any conscious selection. Still it is true that the different pairs of raters did not rate an equal proportion of persons of the different nationalities; in other words, the different sets of raters did not play an equal part in determining the average rating of persons belonging to different nationalities. This introduces a further error which would only be overcome when larger numbers of students are rated.

TABLE 11.

RELATIVE POSITION OF THE DIFFERENT NATIONALITIES IN RESPECT TO MEAN JUDGMENTS ABOUT THEM.

	Leadership	Pertinacity	Humor	Frankness	Suspiciousness	pathy Sym-	Loyalty	Generosity	Obtrusiveness	Coolness
Highest Group	Germans	Germans	Germans Irish Jews	Germans	Jews Irish	Germans	Germans Italians Irish	No significant differences	Jews	Germans Irish
Middle Group	Jews	Jews		Jews	Italians	Jews *Italians*	Jews		Germans *Irish* Italians	Italians Jews
Lowest Group	Irish Italians	Irish Italians	Italians	Irish Italians	Germans	Irish				

TABLE III.

STANDARD DEVIATIONS AND THEIR PROBABLE ERRORS.

Trait.	No. I Germany.	No. 2 Ireland.	No. 3 Italy.	No. 4 Austria.	No. 5 Russia.	No. 6 Jews.	Averages of All Nations Columns 1-5
I. Leadership.	20.0±2.13	27.6±3.52	29.0±2.27	29.5±1.85	26.5±1.69	28.1±1.26	26.5
II. Pertinacity.	20.9±2.23	24.7±3.04	25.0±1.96	23.3±1.45	22.0±1.39	22.7±1.01	23.2
III. Humor.	22.6±2.42	26.4±3.37	26.6±2.12	25.8±1.60	26.7±1.69	26.4±1.17	25.6
IV. Frankness.	18.8±2.00	21.6±2.66	23.6±1.85	22.3±1.38	21.8±1.40	22.2±0.99	21.6
V. Suspiciousness.	21.6±2.31	12.5±1.54	17.4±1.40	20.3±1.27	16.6±1.06	18.6±0.83	17.7
VI. *Sympathy.*	12.0±1.28	24.2±3.08	17.3±1.40	20.9±1.31	16.5±1.05	19.8±0.88	18.2
VII. Loyalty.	15.7±1.67	12.5±1.54	10.2±0.82	17.5±1.09	14.9±0.94	16.4±0.73	14.2
VIII. Generosity.	14.7±1.57	9.7±1.29	18.6±1.48	18.0±1.15	19.9±1.32	19.0±0.87	16.2
IX. Obtrusiveness.	17.4±1.86	18.5±2.28	22.2±1.77	22.3±1.40	21.7±1.37	22.0±0.98	20.4
X. Coolness	15.2±1.63	20.6±2.54	22.4±1.78	22.8±1.41	22.7±1.46	22.8±1.02	20.7
Average of all Traits.	17.9	19.8	21.2	22.3	20.9	21.8	

A CONTRIBUTION TO THE TECHNIQUE OF PARTIAL CORRELATION

By C. L. Huffaker
State University of Iowa

It is well recognized that the partial regression equation is the best tool for the study of many educational problems. This is particularly true if the problem involves (1) the analysis of a "general ability" into its component specific abilities, or (2) the determination of the diagnostic value of different tests. In the schema developed by Yule[1] the computation becomes very complicated and laborious if the number of variables is at all numerous. Methods of shortening the labor have been devised by Kelley[2] and Rosenow[3]. The schema devised by Rosenow requires only about half the labor of that of Kelley, but it has the very important shortcoming that the signs of all but two of the highest order coefficients are indeterminate. It necessarily follows that the signs of all the regression coefficients are indeterminate except two. This shortcoming imposes a serious limitation on the usefulness of Rosenow's schema. Both plans referred to above, make use of the fact, that usually the solution of such problems as have been indicated before requires only one regression equation.

The writer has devised a schema for solving problems in partial regression which involves the finding of fewer coefficients than either of the two plans cited before and at the same time avoids the approximation of Kelley's plan and the indeterminate signs of Rosenow's plan.

NUMBER OF COEFFICIENTS REQUIRED*

Variables	This plan		Yule		Kelley		Rosenow	
	Par.	Mul.	Par.	Mul.	Par.	Mul.	Par.	Mul.
3	2	2	3	1	2	1	2	2
4	9	2	19	1	15	1	9	3
5	22	2	61	1	36	1	22	4
6	46	2	240	0	78	1	45	5

This plan is outlined below. In the outline it is assumed that the reader has a working knowledge of the theory of correlation and of the notations of partial correlations as given by Yule.

The general problem before us is: given a dependent variable and a number of independent variables how can the maximum of information be obtained with the minimum of arithmetic? This problem may be divided into two sub-problems which are successive stages of a single operation. The two stages are: (1) The finding of the coefficients of partial correlation, with reference to the dependent variable, of the highest order; (2) The finding of the coefficients of regression (weights) for estimating the dependent variable.

---

[1] Yule, G. Udny: *An Introduction to the Theory of Statistics*, 1910.

[2] Kelley Truman L.: *Chart to Facilitate the Calculation of Partial Coefficients of Correlation and Regression Equations*, 1921.

[3] Rosenow, Curt.: *The Analysis of Mental Functions*, Psych. Monographs, Vol. XXIV, No. 5, 1917.

*Number of coefficients needed for Kelley's and Rosenow's plans is from Rosenow op.cit.

In a problem involving four variables, the tables given below include all the partial correlation coefficients of all orders needed for finding the coefficients of the highest order with reference to the dependent variable.

First Order Coefficients	May be obtained from Zero Order Coefficients.		
$r_{12.4}$	r_{12}	r_{14}	r_{24}
$r_{13.2}$	r_{13}	r_{12}	r_{23}
$r_{13.4}$	r_{13}	r_{14}	r_{31}
$r_{14.2}$	r_{14}	r_{12}	r_{24}
$r_{23.4}$	r_{23}	r_{24}	r_{34}
$r_{31.2}$	r_{31}	r_{23}	r_{21}

Second Order Coefficients Needed	May be obtained from First Order Coefficients.		
$r_{12.34}$	$r_{12.4}$	$r_{13.4}$	$r_{23.4}$
$r_{13.24}$	$r_{13.4}$	$r_{12.4}$	$r_{23.4}$
$r_{14.23}$	$r_{14.2}$	$r_{13.2}$	$r_{31.2}$

The regression equation to be used in estimating the dependent variable is,

$$\overline{X}_1 = b_{12.34} X_2 + b_{13.24} X_3 + b_{14.23} X_4$$

In which X_1 is the estimated deviation from the mean of variable 1; X_2 observed deviation from the mean of variable 2 etc.

In addition to the coefficients given above the four standard deviations of the type $\sigma 1.234$ will be needed. Each of these may be found in three different ways, for example $\sigma 1.234$ may be found from:

$$\begin{aligned}\sigma^2_{1.234} &= \sigma^2_1(1-r^2_{12})(1-r^2_{13.2})(1-r^2_{14.23})\\ &= \sigma^2_1(1-r^2_{14})(1-r^2_{13.4})(1-r^2_{12.34})\\ &= \sigma^2_1(1-r^2_{13})(1-r^2_{14.3})(1-r^2_{12.34})\end{aligned}$$

It is now possible to select formulae for the standard deviations needed in such a way that all the coefficients used occur in the tables of partial coefficients given above. The formulae selected in this case are:

$$\sigma^2_{1.234} = \sigma^2_1(1-r^2_{12})(1-r^2_{13.2})(1-r^2_{14.23})$$

$$\sigma^2_{2.134} = \sigma^2_2(1-r^2_{24})(1-r^2_{23.4})(1-r^2_{12.34})$$

$$\sigma^2_{3.124} = \sigma^2_3 (1-r^2_{34})(1-r^2_{23.4})(1-r^2_{13.24})$$

$$\sigma^2_{4.123} = \sigma^2_4 (1-r^2_{24})(1-r^2_{34.2})(1-r^2_{14.23})$$

As a check on the arithmetic of the operation, it is well to find multiple R in two different ways, namely by

$$1-R^2_{1(234)} = \sigma\ (1-r^2_{12})(1-r^2_{13.2})(1-r^2_{14.23})$$

$$1-R^2_{1(234)} = \sigma\ (1-r^2_{14})(1-r^2_{13.4})(1-r^2_{12.34})$$

This furnishes an absolute check on all the work except that of $r_{13.24}$ and the coefficients that are used in finding it are checked for their accuracy. There is no check on the arithmetic of the stand deviations included in the plan. However, if a check is desired, they may be found from the following equations, after $r_{24.3}$ has been calculated from r_{24}, r_{23}, and r_{34}.

$$\sigma^2_{1.234} = \sigma^2_1 (1-r^2_{14})(1-r^2_{13.4})(1-r^2_{12.34})$$

$$\sigma^2_{2.134} = \sigma^2_2 (1-r^2_{23})(1-r^2_{24.3})(1-r^2_{12.34})$$

$$\sigma^2_{3.124} = \sigma^2_3 (1-r^2_{23})(1-r^2_{34.2})(1-r^2_{13.24})$$

$$\sigma^2_{4.123} = \sigma^2_4 (1-r^2_{34})(1-r^2_{24.3})(1-r^2_{14.23})$$

In using this plan it should be held in mind that the different partial coefficients are obtained from the column headed "May be obtained from" by substitution in the general formula

$$r_{12.34\,.\,.\,n} = \frac{r_{12.34\,.\,(n-1)} - r_{1n.34\,.\,.\,(n-1)} \cdot r_{2n\,.\,34\,.\,.\,(n-1)}}{\sqrt{1-r^2_{1n.34\,.\,.\,(n-1)}}\ \sqrt{1-r^2_{2n.34\,.\,.\,(n-1)}}}$$

In each case, the first coefficient given becomes the first term of the numerator and the second and third the product-term on the numerator. The second and third coefficients likewise appear in the $1-r^2$ of the denominator. For example:

$$r_{12.4} = \frac{r_{12} - r_{14} \cdot r_{24}}{\sqrt{1-r^2_{14}}\ \sqrt{1-r^2_{24}}}$$

$$r_{12.34} = \frac{r_{12.4} - r_{13.4} \cdot r_{23.4}}{\sqrt{1-r^2_{13.4}}\ \sqrt{1-r^2_{23.4}}}$$

The regression coefficients are obtained by substitution in the general formula

$$b_{12.34..n} = r_{12.34..n}\frac{\sigma_{1.234...n}}{\sigma_{2.134...n}}$$

For example $b_{12.34}$ *and* $b_{13.24}$ *are given by the formulae*

$$b_{12.34} = r_{12.34}\frac{\sigma_{1.234}}{\sigma_{2.134}}$$

$$b_{13.24} = r_{13.24}\frac{\sigma_{1.234}}{\sigma_{3.124}} .$$

For five variables the equation becomes

$$X_1 = b_{12.345}X_2 + b_{13.245}X_3 + b_{14.235}X_4 + b_{15.234}X_5$$

and may be solved by the following outline using the same checks given for four variables, after extending them to include the extra variable.

Fist Order Coefficients Needed	May be obtained from Zero Order Coefficients		
$r_{12.5}$	r_{12}	r_{15}	r_{25}
$r_{13.2}$	r_{13}	r_{12}	r_{23}
$r_{13.5}$	r_{13}	r_{15}	r_{35}
$r_{14.2}$	r_{14}	r_{12}	r_{24}
$r_{14.5}$	r_{14}	r_{15}	r_{45}
$r_{15.2}$	r_{15}	r_{12}	r_{25}
$r_{23.5}$	r_{23}	r_{25}	r_{35}
$r_{24.5}$	r_{24}	r_{25}	r_{45}
$r_{34.2}$	r_{34}	r_{23}	r_{24}
$r_{34.5}$	r_{34}	r_{35}	r_{45}
$r_{35.2}$	r_{35}	r_{23}	r_{25}
$r_{45.2}$	r_{45}	r_{24}	r_{25}

Second Order Coefficients Needed	May be obtained from First Order Coefficients		
$r_{12.45}$	$r_{12.5}$	$r_{14.5}$	$r_{24.5}$
$r_{13.45}$	$r_{13.5}$	$r_{14.5}$	$r_{34.5}$
$r_{14.23}$	$r_{14.2}$	$r_{13.2}$	$r_{34.2}$
$r_{15.23}$	$r_{15.2}$	$r_{13.2}$	$r_{35.2}$
$r_{23.45}$	$r_{23.5}$	$r_{24.5}$	$r_{34.5}$
$r_{45.23}$	$r_{45.2}$	$r_{35.2}$	$r_{34.2}$

Third Order Coefficients Needed	May be obtained from Second Order Coefficients		
$r_{12.345}$	$r_{12.45}$	$r_{13.45}$	$r_{23.45}$
$r_{13.245}$	$r_{13.45}$	$r_{12.45}$	$r_{23.45}$
$r_{14.235}$	$r_{14.23}$	$r_{15.23}$	$r_{45.23}$
$r_{15.234}$	$r_{15.23}$	$r_{14.23}$	$r_{45.23}$

The formulae for the standard deviations are:

$$\sigma^2_{1.2345}=\sigma^2_1(1-r^2_{12})(1-r^2_{13.2})(1-r^2_{14.23})(1-r^2_{15.234})$$

$$\sigma^2_{2.1345}=\sigma^2_2(1-r^2_{25})(1-r^2_{24.5})(1-r^2_{23.45})(1-r^2_{12,345})$$

$$\sigma^2_{3.1245}=\sigma^2_3(1-r^2_{35})(1-r^2_{34.5})(1-r^2_{23.45})(1-r^2_{13.245})$$

$$\sigma^2_{4.1235}=\sigma^2_4(1-r^2_{24})(1-r^2_{34.2})(1-r^2_{45.23})(1-r^2_{14.235})$$

$$\sigma^2_{5.1234}=\sigma^2_5(1-r^2_{25})(1-r^2_{35.2})(1-r^2_{45.23})(1-r^2_{15.234})$$

Formulae for the b's are

$$b_{12.345}=r_{12.345}\frac{\sigma 1.2345}{\sigma 2.1345}$$

$$b_{13.245}=r_{13.245}\frac{\sigma 1.2345}{\sigma 3.1245}$$

etc.

The equation for six variables is:

$$X_1 = b_{12.3456} X_2 + b_{13.2456} X_3 + b_{14.2356} X_4 + b_{15.2346} X_5 + b_{16.2345} X_6$$

The plan for six variables is given below and is subject to the same checks as the problems with a smaller number of variables.

Fist Order Coefficients Needed	May be obtained from Zero Order Coefficients		
$r_{12.6}$	r_{12}	r_{16}	r_{26}
$r_{13.2}$	r_{13}	r_{12}	r_{23}
$r_{13.6}$	r_{13}	r_{16}	r_{36}
$r_{14.2}$	r_{14}	r_{12}	r_{24}
$r_{14.6}$	r_{14}	r_{16}	r_{46}
$r_{15.2}$	r_{15}	r_{12}	r_{25}
$r_{15.6}$	r_{15}	r_{16}	r_{56}
$r_{16.2}$	r_{16}	r_{12}	r_{26}
$r_{23.6}$	r_{23}	r_{26}	r_{36}
$r_{24.6}$	r_{24}	r_{26}	r_{46}
$r_{25.6}$	r_{25}	r_{26}	r_{56}
$r_{34.2}$	r_{34}	r_{23}	r_{24}
$r_{34.6}$	r_{34}	r_{36}	r_{46}
$r_{35.2}$	r_{35}	r_{23}	r_{25}
$r_{35.6}$	r_{35}	r_{36}	r_{56}
$r_{36.2}$	r_{36}	r_{23}	r_{26}
$r_{45.2}$	r_{45}	r_{24}	r_{25}
$r_{45.6}$	r_{45}	r_{46}	r_{56}
$r_{46.2}$	r_{46}	r_{24}	r_{26}
$r_{56.2}$	r_{56}	r_{25}	r_{26}

Second Order Coefficients Needed	May be obtained from First Order Coefficients		
$r_{12.56}$	$r_{12.6}$	$r_{15.6}$	$r_{25.6}$
$r_{13.56}$	$r_{13.6}$	$r_{15.6}$	$r_{35.6}$

$r_{14.23}$	$r_{14.2}$	$r_{13.2}$	$r_{34.2}$
$r_{14.56}$	$r_{14.6}$	$r_{15.6}$	$r_{45.6}$
$r_{15.23}$	$r_{15.2}$	$r_{13.2}$	$r_{35.2}$
$r_{16.23}$	$r_{16.2}$	$r_{13.2}$	$r_{36.2}$
$r_{23.56}$	$r_{23.6}$	$r_{25.6}$	$r_{35.6}$
$r_{24.56}$	$r_{24.6}$	$r_{25.6}$	$r_{45.6}$
$r_{34.56}$	$r_{34.6}$	$r_{45.6}$	$r_{35.6}$
$r_{45.23}$	$r_{45.2}$	$r_{35.2}$	$r_{34.2}$
$r_{46.23}$	$r_{46.2}$	$r_{34.2}$	$r_{36.2}$
$r_{56.23}$	$r_{56.2}$	$r_{35.2}$	$r_{36.2}$
Third Order Coefficients Needed	May be obtained from Second Order Coefficients		
$r_{12.356}$	$r_{12.56}$	$r_{13.56}$	$r_{23.56}$
$r_{12.456}$	$r_{12.56}$	$r_{14.56}$	$r_{24.56}$
$r_{13.456}$	$r_{13.56}$	$r_{14.56}$	$r_{34.56}$
$r_{14.356}$	$r_{14.56}$	$r_{13.56}$	$r_{34.56}$
$r_{15.234}$	$r_{15.23}$	$r_{14.23}$	$r_{45.23}$
$r_{16.234}$	$r_{16.23}$	$r_{14.23}$	$r_{46.23}$
$r_{23.456}$	$r_{23.56}$	$r_{24.56}$	$r_{34.56}$
$r_{24.356}$	$r_{24.56}$	$r_{23.56}$	$r_{34.56}$
$r_{56.234}$	$r_{56.23}$	$r_{45.23}$	$r_{46.23}$
Fourth Order Coefficients Needed	May be obtained from Third Order Coefficients		
$r_{12.3456}$	$r_{12.456}$	$r_{13.456}$	$r_{23.456}$
$r_{13.2456}$	$r_{13.456}$	$r_{12.456}$	$r_{23.456}$
$r_{14.2356}$	$r_{14.356}$	$r_{12.356}$	$r_{24.356}$
$r_{15.2346}$	$r_{15.234}$	$r_{16.234}$	$r_{56.234}$
$r_{16.2345}$	$r_{16.234}$	$r_{15.234}$	$r_{56.234}$

Formulae for standard deviations,

$$\sigma^2_{1.23456}=\sigma^2_1(1-r^2_{16})(1-r^2_{15.6})(1-r^2_{14.56})(1-r^2_{13.456})(1-r^2_{12.3456})$$

$$\sigma^2_{2.13456}=\sigma^2_2(1-r^2_{26})(1-r^2_{25.6})(1-r^2_{24.56})(1-r^2_{23.456})(1-r^2_{12.3456})\;.$$

$$\sigma^2_{3.12456}=\sigma^2_3(1-r^2_{36})(1-r^2_{35.6})(1-r^2_{34.56})(1-r^2_{23.456})(1-r^2_{13.2456})$$

$$\sigma^2_{4.12356}=\sigma^2_4(1-r^2_{46})(1-r^2_{45.6})(1-r^2_{34.56})(1-r^2_{24.356})(1-r^2_{14.2356})$$

$$\sigma^2_{5.12346}=\sigma^2_5(1-r^2_{25})(1-r^2_{35.2})(1-r^2_{45.23})(1-r^2_{56.234})(1-r^2_{15.2346})$$

$$\sigma^2_{6.12345}=\sigma^2_6(1-r^2_{26})(1-r^2_{36.2})(1-r^2_{46.23})(1-r^2_{56.234})(1-r^2_{16.2345})$$

Formulae for the b's are:

$$b_{12.3456}=r_{12.3456}\frac{\sigma 1.23456}{\sigma 2.13456}$$

$$b_{13.2456}=r_{13.2456}\frac{\sigma 1.23456}{\sigma 3.12456}$$

etc.

The same plan of selection can be used for any number of variables with the same proportionate saving of labor. The use of Kelley's Chart referred to before is possible with four variables, but is not recommended for more than that number. For five and six variables the writer recommends the use of such tables as Miner's[1] and a calculating machine. By the use of these tables and a calculating machine it is possible to solve a problem of six variables correct to four places in three hours, after the Zero Order Coefficients have been found.

---

[1] Miner, *Tables of* $1-r^2$ *and* $\sqrt{1-r^2}$ John Hopkins Press, 1922.

THE RELATION BETWEEN THE INTELLIGENCE AND VOCATIONAL CHOICES OF HIGH SCHOOL PUPILS

By Gustave A. Feingold
Hartford Public High School.

In the varied literature on vocational guidance and vocational education, one finds that almost nothing has been done in the attempt to correlate the vocational choices of high school pupils with their mentality.[1] Yet if vocational guidance is to be put on a scientific as well as practical basis, this is the very first thing that should be done. The chief function of the vocational adviser in the high school should be to save pupils of high mental ability from entering occupations of limited opportunity, on the one hand, and to advise others of little ability against trying to enter occupations for which they have not the intellectual foundation, on the other hand.

In the September 2, 1922, issue of School and Society there appears an article on Occupational-Intelligence-Standards[2] wherein is presented an intelligence scale grouping some ninety-six different occupations into various Army Alpha scores, based on the results of the occupational studies made by the Division of Psychology, S. G. O.; and reported in the now famous Volume XV, Memoirs of National Academy of Sciences.

With this Occupational-Intelligence Scale before us, there was at hand a means for comparing the vocational choices of high school pupils with their intelligence grades, in order to determine whether they overrated or underrated their abilities, and to what extent.

All occupational-intelligence scales are based on the assumption that certain vocations are within reach of certain mental levels. One group of vocations is within reach of "A" intelligence, another within reach of "C" intelligence, and so forth. If a high school pupil of a certain mental level, "D," let us say, desires to enter a vocation which belongs to a much higher level, like "B", he overestimates his ability by two degrees. The assumption is that he will either not succeed in entering his desired vocation, or if he does, the probability is that he will not be efficient in it. On the other hand, if a boy expects to enter an occupation that belongs to a much lower mental level than his own, the community loses by that choice, inasmuch as it does not get from him the best and utmost service that he is capable of rendering, nor will he be happy because he will not have the full opportunity to express his entire personality in it. This is an important matter, for about 50 per cent of all high school pupils make up their minds with respect to their future calling even before they enter school, and about 80 per cent of these cling to their original choice throughout their school career. Our problem, therefore, was to find out to what extent high school pupils make or express vocational choices that are within reach of their mentality.

---

[1] In the 183 titles on *The Psychology of Vocational Selection* reviewed by A. W. Krohnhauser, Psy. Bull. Vol. 19, No. 4, 1922, not one seems to deal specifically with the above problem.

[2] Douglas Fryer, *Occupational-Intelligence Standards*, School and Society, Sept. 2, 1922, p. 270.

To be sure, the alpha scores given for the various occupational groups in Fryer's table were attained by individuals whose schooling, for the most part, lay from five to fifteen years behind them; and, in view of the fact that when the youth gets out of school he is bound, more or less, to get into a rut of thinking and acting, it is inevitable that he should strike a lower score in an intelligence test then, than he would have done when he was at the height of his school days.[3] This is demonstrated by the fact that in the relative mental ranks of the various occupational groups in the afore-mentioned Volume XV, the "student" is placed third from the top, apparently above the physician, above the clergyman, and above the accountant, notwithstanding the fact that all of these have been students at one time in their lives.[4] Consequently, a certain amount of caution had to be exercised in comparing the intelligence grades of high school individuals with those of mature persons in various occupational groups. For, a high school senior may readily score 127 points, which is the average for the physician group on the Occupational-Intelligence Scale referred to, and yet be woefully deficient in mentality for the medical profession. In other words, his relatively high score may be due to the educational increment, which is bound to innure to high school pupils of every grade of mentality by virtue of their constant immersion, forced or otherwise, in the educational waters.[5] It was, therefore, necessary to strike a group of high school pupils not too long domiciled in the realm of secondary education, on the one hand, and among whom there would be found a mental level that would correspond, in points scored, to the mentality of some definite occupational group in Fryer's scale, on the other hand.

---

[3] Since the above was written Margaret V. Cobb's *The Limits Set to Educational Achievement by Limited Intelligence*, Jour. Educ. Psy. Vol. XIII No. 6, 1922, has made its appearance and from the figures presented in Table VIII, one may infer that those who have had only one year of high school education, will, 15 years later, score seven points more than they did as high school freshmen, those who left high school at the end of the sophomore year will, 15 years later, score five points less than as sophomores, adults of three years' high school training will score 10 less than they did as juniors and those of four years high school will score 20 points less than they did as high school seniors.

[4] Mem. Nat. Ac Sc. Vol. XV, p. 830.

[5] For substantiation of this see Cobb Op. cit. p. 402-403.

Now it happened that in June, 1922, all prospective Freshmen of the Hartford Public High School, some 1200 in number, were given a modified Army Alpha test and the passing mark was set at 70 points, equivalent to an I. Q. of 103 and to 50 per cent intelligence on our scale of 100. By dividing that scale into ten deciles, the uppermost, or "A" mental level, corresponded exactly with the occupational groups that fell into the "A" mental level on Fryer's scale. Thereafter, it was merely necessary to draw lines through all corresponding points of division on the Fryer scale and note whether the vocational choice made by a pupil of a given degree of intelligence fell within his mental level on the converted scale or not.

In one respect Fryer's Occupational-Intelligence Scale is unsatisfactory for high school purposes in that it is not sufficiently complete in the matter of vocations. No provision is made in it for the professions of law, banking, journalism, high school and college teaching, nor for teachers of music, gymnastics, and cooking. Yet, there must have been hundreds of lawyers, high school teachers, college professors, bank tellers, not to mention instructors of gymnastics, enrolled in the army. In an intelligence test given to fifty-five high school teachers, the writer found that their median on the Army Alpha test was 162 points, with a variation that ranged from 94 to 190. Also, in the occupational studies made by the Division of Psychology, S. G. O., lawyers are ranked second from the top. The writer felt justified, therefore, in placing lawyers, journalists, high school and college teachers in the "A" class on Mr. Fryer's scale. Authors and bankers were placed in the modified "B" group. Music, gymnasium, elocution, and cooking teachers, professions selected by high school pupils, as well as the indefinite profession "business", so frequently chosen by high school boys and girls, were placed in the modified "D" group. Among girls, those who chose dressmaking as their vocational ambition, were classified with tailors, and hair-dressers with barbers.

Investigations conducted by various psychologists have demonstrated that the I. Q. does not change appreciably. The writer has thus far retested two high school classes at intervals of two years and has found practically no change in the average I. Q. attained by the same pupils as Freshmen, and two years later as Juniors. Thus, the average I. Q. of 288 Freshmen who had entered the Hartford High School in September, 1920, was 111[6]; their average age having been 14 years and 6 months, and the average number of points scored by them on the writer's modified Alpha test having been 96, which is equivalent to a mental age of 16 years and 2 months. The average number of points scored by these same 288 pupils in September, 1922, when they became Juniors was 116, which is equivalent to the mental age of 17 years and 1 month; their average I. Q. was consequently 104."

---

[6] Their educational mental increment was about 5 Alpha points. Deducting that number from 110 gives them an I. Q of 102. This harmonizes with theory—as the normal individual approaches adult mentality his I. Q. should approach 100.

By adopting a percentage scale of mental alertness, the base of which is 152 points, equivalent to 100 per cent intelligence, for high school Freshmen tested three weeks before graduation from grammar school, and a scale with a base of 172 points for Sophomores, one with a base of 186 points for Juniors, and one with a base of 196 points for Seniors, the latter three groups tested at the beginning of their class careers, the writer has found but little displacement from one mental level to the other as the pupils progressed from the Freshman to the Senior year and were subjected to two separate tests.

Thus, the average per cent of intelligence for the above mentioned 288 pupils, rated by their Freshman scores, was 59.0, and their average intelligence two years afterwards, rated by their Junior scores, was 62.3. The coefficient of correlation between the Freshman and Junior scores was +.875±.009. Fifty, on a scale of 100, being the minimum requirement for passing our high school intelligence test, the corresponding passing I. Q.'s for the Freshman, Sophomore, Junior, and Senior classes are, respectively, 103, 101, 98, and 95. The following table gives the points scored, per cent intelligence, and letter grade for each mental level in the four high school classes.

TABLE I.

COMPARATIVE INTELLIGENCE SCALES FOR HIGH SCHOOL *CLASSES*.

(To be applied to the last three at the beginning of their class careers)

POINTS SCORED					
Freshmen	Sophomores	Juniors	Seniors	% Int.	Letter Grade
137-152	165-172	167-186	176-196	90-100	A
121-136	138-154	148-166	156-175	80-89	B
106-120	120-137	130-147	136-155	70-79	C
91-105	103-119	111-129	116-135	60-69	D
76-90	86-102	92-110	96-115	50-59	E
61-75	69-85	74-91	76-95	40-49	F
45-60	58-68	55-73	56-75	30-39	G
44 & Less	51 & Less	54 & Less	55 & Less	29 & Less	H

If now we take Fryer's Occupational-Intelligence Scale, modified so as to include the professions frequently chosen by high school pupils, and then group into the various mental grades the professions that naturally fall into them by virtue of the average points scored by the members in each group, we obtain the following Occupational-Intelligence Scale especially adapted for high school pupils, and, though based on the intelligence achievements of high school Freshmen, is none the less applicable to upper classmen, when attention is paid to the letter grade in mentality instead of points scored.

TABLE II.

OCCUPATIONAL-INTELLIGENCE SCALE FOR HIGH SCHOOL PUPILS.

Letter Grade	Av. I. Q.[1]	Av. No. Points		Occupation
A	126	162	141	Teacher (College & High)
		161		Engineer (Civil & Mech.)
		?		Lawyer
		?		Journalist
		152		Clergyman
		137		Accountant
B	122	127	126	Physician
		?		Novelist
		122		Teacher (Grades)
		?		Banker
		119		Chemist
C	116	114	110	Draftsman
		111		Secretary (Private & Institutional)
		110		Dentist
		109		Executive, minor
D	110	103	95	Stenographer & Typist
		101		Bookkeeper
		99		Nurse
		96		Clerk (office)
		91		Clerk (railroad)
		?		"Business"
		?		Gymnasium Teacher
		?		Cooking Teacher
		?		Music Teacher
		?		Elocution Teacher
E	103	86	80	Photographer
		85		Telegrapher & Radio Operator
		82		Musician (Band & Orchestra)
		81		Artist (Sign Letterer)
		81		Clerk (Postal)
		81		Electrician
		80		Foreman (Construction)
		78		Druggist
		77		Foreman (Factory)

---

[1] Derived from the Freshmen scores and ages in each group.

		70		Telephone Operator
		69		Carpenter (Ship)
		69		*Policeman & Detective*
		68		Auto Assembler
		67		Tool Maker
		66		Plumber
F	95	65	65	Lath Hand (Production)
		65		Auto Mechanic (General)
		65		Auto Chauffeur
		65		Tailor
		?		Dressmaker
		?		Milliner
		63		Machinist (General)
		62		Actor (Vaudeville)
		60		Printer
		60		*Carpenter* (General)
		59		Painter
		58		Farmer
G	90	58	55	Brick Layer
		57		Caterer
		55		Barber
		?		Hair Dresser
		52		Sales Clerk
		40		Hospital Attendent
H	84	32	35	Sailor
		31		Structural Steel Worker

Not all of the occupations given in Fryer's scale are included in this modified one. Only those which were chosen at least once among the 512 pupils who expressed their vocational preference are given a place in it. These are printed in heavy type. The remaining occupations are included as ones which the writer thought high school pupils might possibly choose, at least temporarily; also in order to serve as a means of comparison between the different types of vocational mental levels.

Let us now see what the 512 high school Freshmen actually did in the matter of choosing a vocation or indicating a vocational preference. Table III shows the degrees and the direction of variation of vocational choice from mental level among the 250 boys. The numbers +5, +4, +3, etc., represent the number of mental levels, above that of the individual, wherein lay the vocation he chose. Minus one, —2, etc., indicate the number of mental levels below that of the individual wherein lay the profession he selected. The letters in the squares show in what mental level lay the vocation chosen; the numbers in the squares represent the number of pupils making that choice.

The above table shows some astonishing results. There is a conspicuous tendency for the superior pupils, those of "A" and "B" mentality, to underrate their ability and to choose a profession that falls in a mental level considerably below their own. And there is a still greater tendency for pupils of low mentality to choose professions that lie considerably above their mental levels. To be sure, the choosing of a profession that lies within one mental level above or below the one to which a particular individual belongs is no serious mistake. It is perfectly safe to assume that an individual of "B" mentality is quite able to pursue a profession that lies in the "A" mental level, provided he puts forth enough industry. Likewise, there is no reason to suppose that a pupil of "C" mentality is throwing his future away entirely merely by choosing a profession that belongs to the "D" mental level. Therefore, in the treatment of the data presented in Tables III and IV we shall consider all individuals who vary but one degree in either direction as having made a proper vocational choice. It is only those whose vocational choice lies within two or more mental levels above or below their own who will be considered as not having chosen wisely.

With this proviso before us we find that of the thirty superior boys only 20, or 66.6 per cent, make a proper choice, whereas 33.3 per cent choose some profession that requires far less ability than they possess, as judged by the intelligence tests. Of course, for the "A" pupils, there was no alternative but to choose a profession that lay either within their mental level or in some lower level; but why should three of the four "A" boys choose vocations that are so far below their own mental grade? The "B" pupils seem to show somewhat more courage, as well as more wisdom, than the "A" individuals, in choosing their vocations; inasmuch as 73 per cent of them select vocations that fall within their mental capacity. Among the boys of "C" ability only 44 per cent seem to make a proper choice, while the number that overrate their ability is twice as large as those who underrate it. When we come to those who failed in the intelligence tests, having scored a grade of "F" and lower, which is equivalent to 76 points and less on the Army Alpha scale, we find that no less than 88 per cent of them select professions that are anywhere from two to five mental levels above their own.

To be sure the claim may be made, and justly too, that the so-called professions lie within the three highest mental levels, and inasmuch as the very presence of boys in high school already indicates a leaning toward some profession, it is only natural that *they should* tend to overrate their ability, as judged by their vocational preference, rather than to underrate it. But this simply *shifts the argument back one step by taking the problem out of* the high school and placing it within the grammar school and home. *It merely means that American children tend to overrate their* ability even before they get into high school. Judged by the large proportion of high school failures, it being something like 60 per cent, one must say that the above statement is not far from being accurate. Moreover it is significant to note that among the mental failures a larger proportion, fully 48 per cent, choose a profession that lies in the "A" and "B" levels than do the pupils of either "E" or "D" ability, the percentages of excessive vocational overreaching in their cases being 40 and 42, respectively. In other words, the distribution table plainly shows that the lower an individual stands in mentality, the higher seems to be his vocational ambition. Whether this is a manifestation of compensation, the desire filling the gap within the mind, or whether it is simply a further indication of poor judgment and inability to envisage the future in matters practical as well as ideal, is a topic for separate discussion.

Nevertheless, there is sufficient evidence at hand to indicate that it is the latter and not entirely the former. Investigating the proportion of pupils in the various mental levels who survived the *high school curriculum for at least two years, the writer has* found that in the Hartford Public High School fully 73 per cent *of those who fail in intelligence at the beginning of their Freshman* year do not succeed in becoming Juniors. Either they drop out of school altogether, or remain in a lower class because of poor scholarship. According to Cobb,[8] 87 per cent of the Freshmen who score below 77.5 points drop out before the Senior year in the Michigan high schools, and in Madsen's group about 84 per cent of such individuals drop out. Now, in view of the fact that the professions within the "A" and "B" groups require considerably more than mere high school training, it is quite obvious that of the thirty-seven "F" and "G" mental level pupils who selected professional occupations for their career, practically three-quarters of them will not proceed far enough in higher training even to come within reach of a high school diploma. Once more according to Cobb, only about 4.5 per cent of them may be expected to enter college.[9]

---

[8] *The Limit Set to Educational Achievement by Limited Intelligence*, Jour. Edu. Psy., Vol. XVIII, No. 3, p. 164.

[9] Op. cit. Table IX, p. 169.

TABLE III.

DEGREE AND DIRECTION OF VARIATION OF VOCATIONAL CHOICE FROM MENTAL LEVEL AMONG BOYS.

Degrees of Variation	+5	+4	+3	+2	+1	0	-1	-2	-3	-4	-5	Total	Per Cent
Mental Level													
A						1 A	B	C	2 D	E	1 F	4	1.6
B					14 A	5 B	C	3 D	1 E	2 F	1 G	26	10.4
C				14 A	7 B	2 C	8 D	2 E	5 F			38	15.2
D			21 A	8 B	3 C	26 D	7 E	4 F	1 G			69	27.6
E		15 A	5 B	C	20 D	4 E	4 F					48	19.2
F	20 A	5 B	6 C	12 D	3 E	1 F	2 G					49	19.6
G	6 B	C	5 D	2 E	1 F	1 G						15	6.0
H			1 E	F	G	H						1	0.4
Total	26	20	38	36	48	39	21	9	9	2	2	250	
Per Cent	10.4	8.0	15.2	14.4	19.2	15.6	8.4	3.6	3.6	0.8	0.8		100.0

48.0% 43.2% 8.8%

TABLE IV.

DEGREE AND DIRECTION OF VARIATION OF VOCATIONAL CHOICE FROM MENTAL LEVEL AMONG GIRLS

Degrees of Variation	+ 5	+ 4	+ 3	+ 2	+1	0	-1	-2	-3	-4	-5	Total	Per Cent
Mental Level													
A						A	2 B	1 C	2 D	E	1 F	6	2.3
B					1 A	9 B	1 C	6 D	E	1 F		18	6.9
C				2 A	13 B	2 C	14 D	1 E	1 F			33	12.6
D			1 A	18 B	1 C	34 D	3 E	2 F	1 G			60	22.0
E		4 A	17 B	C	42 D	2 E	1 F					66	25.2
F		14 B	C	42 D	E	F						56	21.4
G	4 B	C	14 D	E	1 F	G						19	7.3
H	1 C	3 D	E	F	G	H						4	1.5
Total	5	21	32	62	58	47	21	10	4	1	1	262	
Per Cent	1.9	8.0	12.2	23.7	22.1	18.0	8.0	3.81	1.5	0.4	0.4		100

45.8% 48.1% 6.1%

To judge by the professions these 250 boys wish to enter, we may say that ten per cent overrate their ability by five mental levels, eight per cent overrate themselves by four mental levels, 15 per cent by three mental levels, and 14 per cent by two mental levels. In brief, 48 per cent of all the boys overrate their ability by choosing a profession that lies in a mental level that is from two to five grades above their own. Forty-three and two tenths per cent seem to choose wisely, in that the profession they select lies within a mental level that varies only one degree from their own. Eight and eight tenths per cent underrate their ability by choosing a vocation that lies in a mental level which is two or more grades below theirs. The coefficient of correlation between the intelligence and vocational choices of these 250 boys is +.045 ±.042. *Practically zero.*

Fryer's Occupational-Intelligence-Scale does not lend itself very readily for the classification of girls in the same manner as boys, for the very good reason that the occupations considered are chiefly those for men. Nevertheless, it contains the professions of teacher, stenographer and typist, bookkeeper, nurse, and clerk, which constitute nearly 90 per cent of the professions selected by high school girls. By including the professions of music, cooking and gymnasium teachers with the class of bookkeepers, nurses and stenographers, as we have done, we have, it seems, made provision for practically every profession that a high school girl would be likely to choose. With these additions, *we were surprised to discover that of the 262 girls, who indicated their professional choice, only one selected a vocation for which no provision could be made on this modified Occupational Scale, and that one was acting, presumably in drama.* Tabulating the girls in the same manner as we did the boys, we obtain Table IV, which shows the extent and the direction in which their vocational choice varies from their mental level.

The similarity in the distribution curve for boys and girls is remarkable. Here, too, we find that 45.8 per cent of the girls overrate their ability by choosing vocations that lie within two or more mental levels above their own. Forty-eight and one tenth per cent make a proper choice and 6.1 per cent underrate themselves. The coefficient of correlation between intelligence and vocational choice for the girls is +.0987 ±.041. In spite of the fact that the scale was based on vocations pursued by men, we find that it is applicable to girls, and, indeed, girls seem to choose their vocations a little more intelligently than do the boys.

Before summing up the results of our study, we must once more forestall the probable objection that the reason why these pupils tend to overrate themselves by choosing vocations that lie so high above their own mental levels is because their intelligence has been judged too early in the stage of mental development. There is no ground for such criticism. First, because the average mental age of these pupils was 15 years and 6 months their average I. Q. 107. This is considerably higher than the average mental age and I. Q. for the American population as a whole. In the second place, the "A" mental group of these pupils coincided exactly with the "A" group on the Occupational-Intelligence Scale in the number of Alpha points scored. None of the deciles in the higher classes could be so compared with the mental groups on Fryer's Occupational-Intelligence Scale. Thirdly, an intelligence test given to 105 evening high school pupils, all of whom were grammar school graduates, some of them being college graduates, some grammar school teachers, one professional nurse, three machinists, and most of the others time-keepers, bookkeepers, store clerks, and office clerks yielded an average score of 90 points, which is only five points above the average for the 512 high school Freshmen considered in the above tables. Fourthly, the average intelligence of our "D" group (95 points) is about equal to that of adults who had one year of high school training and were tested 15 years later.[10] The average of our "C" group is equal to that of adults who had gone through Junior year in high school and were tested 15 years out of school. The average intelligence of our "B" group is higher than the average intelligence of adults who had been graduated from the high school or who had been college freshmen and were tested 15 years later. With respect to the "E" pupils, investigations have thus far shown that not more than 30 per cent of them succeed in becoming high school seniors.

CONCLUSIONS.

The above study leads to the following conclusions:

First: Only about 46 per cent high school pupils make proper vocational choices—choices that can be realized by them and which they will pursue with maximum satisfaction and efficiency.

Second: About the same number, 47 per cent, make vocational choices that lie considerably beyond their mental reach. Presumably, this proportion of pupils will either not gain their desired vocations or, such as do, will not be entirely efficient in them.

Third: Approximately 7 per cent underrate their ability by choosing vocations that lie considerably below their own mental level.

Fourth: It would seem advisable, therefore, that the first duty of the high school vocational guide or vocational educator should be to acquaint himself with the intelligence of the pupils to whom he would give vocational advice.

Fifth: Viewed from the social standpoint it cannot be said that the American youth is devoid of ambition. He may not always put forth the necessary effort to realize his aim, but, judged by the goal that he sets for himself, he certainly has far reaching desires.

---

[10] Cobb, Op. cit. Table 8, p. 400.

RELATION OF MENTAL ALERTNESS TEST SCORE TO POSITIONS AND PERMANENCY IN COMPANY

By Marian A. Bills
Bureau of Personnel Research.
Carnegie Institute of Technology.

Three years ago the Bureau of Personnel Research Carnegie Institute of Technology, gave a mental alertness test to a large clerical force. The total results were made practical use of by the firm. The experiment described below is simply a by-product of that investigation but holds, we believe, several points of interest.

From the test scores of the entire clerical group those of a hundred thirty-three individuals working at five distinct jobs, were drawn off for special study. These five jobs were chosen for three reasons. (1) The jobs were all individual jobs, none of them involving supervisory duties. (2) They represented five distinct levels of difficulty of work. (3) A sufficiently large number of individuals were working at each of the five jobs, so that an average would have meaning.

Thirty months later the subsequent history of those one hundred and thirty-three individuals was studied, as regards one fact, were they still on the job? This gave us a perfectly objective criterion.[1]

The date thus obtained was studied with the following questions in mind.

(1) Is there a relation between positions in the company and mental alertness score? (2) Does a natural selection take place with time that makes this relation more evident?

We would expect natural selection to take place in two ways. First by the persons making low scores in the mental alertness test being unable to handle high grade work and leaving and second by persons making high scores in the mental alertness test becoming dissatisfied with low grade work and leaving. That is at the end of two and a half years, we would expect that the average mental alertness score of those remaining in the high grade jobs would be higher than the average of the original group. And that the average score of those remaining in the low grade jobs would be lower than that of the original group. In short, if mental alertness is a prominent factor in clerical work, we would expect the groups to become more distinct when only those persons are considered that remain in the group more than two and a half years.

---

[1] In this company no definite promotional or rating system had been worked out and promotions or ratings would have been too indefinite a criterion to have been used.

RESULTS.

The correlation between the mental alertness score and the level of difficulty of work performed by the individual was for the original group +.22, for the group remaining after two and one-half years +.41. That in thirty months raised the correlation 19 points.

In order to study the effect of natural selection in more detail, we have split each grade into three classes: (1) Those scoring above 110; (2) Those scoring between 80-110; (3) Those scoring below 80. The dividing scores chosen are arbitrary but were taken because they seemed from the distribution of cases to be natural breaking points. Interest centers around groups one and three. Group two contains about 25 per cent of the cases and may be considered a neutral group. We have lettered the five jobs from A to E, A being the least difficult job, B being more difficult than A, and less difficult than C, and E being the most difficult job.

The following two tables give the data for (1) the original group and (2) the group remaining after two and one-half years.

ORIGINAL GROUP.

Grade of Work	No. in Grade	Median Score	Per cent scoring over 110	Per cent scoring under 80
A	16	85	26	50
B	15	77.5	20	53
C	24	105	45	33
D	56	103.5	46	30
E	22	118.5	50	13

GROUP REMAINING THIRTY MONTHS.

Grade of Work	No. in Grade	Median Scores	Per cent scoring over 110	Per cent scoring under 80	Per cent turnover for those scoring over 110	Per Cent turnover for those scoring under 80
A	7	67.5	0	57	100	37
B	6	80.0	0	50	100	62
C	11	95.0	27	36	72	50
D	26	111.0	51	23	53	58
E	13	123.0	57	7	41	66

The following points would seem to be noteworthy.

(1) In the original group the median scores of the grades range from 77.5 to 118.5, forty-one points. There are two instances where a lower grade job has a slightly higher median score than the next higher grade job. In the group remaining thirty months the median scores of the grades range from 67.5 to 123.0, 55.5 points. Each grade shows a higher median than the grade below it.

(2) In the two least difficult jobs A and B, no one scoring over 110 has remained. In the C grade a smaller per cent of the remaining than the original group have scores over 110, while for the D and E grade, a larger per cent of the remaining group have scores over 110 than of the original group.

(3) Almost the opposite holds when we consider the group scoring below 80. There is a steady decrease of the per cent scoring below 80 with the rise of the grade of work. There is also with one exception an increase in turnover in this group from A to E.

These results are as a whole confirmed by a study of 330 other cases from the same firm where many jobs were lumped together as being of the same difficulty and the grading was much less carefully done. On account, however, of the inaccuracies that we know entered into the grading, we do not believe the results are worth quoting in detail.

A STUDY OF THE CORRELATION OF COLLEGE STUDENTS' ESTIMATES OF INTELLIGENCE WITH THE OTIS TESTS AND OTHER SCALES

By J. U. Yarbrough
Carnegie Institute of Technology, Pittsburg, Pa.

It is a matter of common observation that college students select certain of their number to designate as superior, bright, and intelligent; and just as frequently do they select others from their group and call them inferior, stupid, and dull. Each student soon after entering college is thus tacitly assigned to his group by his fellow students, and thereafter student conduct toward him is in accordance with this assignment, regardless of the method through which or the circumstances under which the judgment was formed. After observing this practice of more or less unavowed placement by students in some of our largest universities, and finding it to be as common as in our smaller colleges; and too, after observing the certainty with which a group of students act toward a fellow student when once it has placed him intellectually, the writer set about to devise some scheme for testing the accuracy with which college students place their fellows in the scale of intelligence.

In order to make this test the judgments of a group of college students were secured and analyzed. The criteria used by each student in estimating the ability of fellow students were listed in terms of human traits. These traits were used as a basis for rating all members of the class. This final rating of the students by the class was correlated with the rating on the basis of their score on the Otis Test, instructor's estimate, and college grades.

METHOD OF ANALYZING THE STUDENT'S JUDGMENT.

The students used in this experiment were of junior and senior rank. They were in a small compact college where the students were brought together daily in their school routine; where, in fact, the success of the existing system of student self government depended largely upon this close contact. Moreover, they had been associated in this class in psychology for a period of six months and had full opportunity to observe each other in class discussions and work in general. Their work in this class had been in general psychology with no special emphasis placed upon the topic of intelligence.

These thirty students were provided with paper and pencil and without any foreknowledge of the content or purpose of the experiment were asked to answer the two questions given below as the experimenter dictated them:

1. Do you make it a habit to rate one of your fellow students, A, as intelligent and another student, B, as stupid?
2. Enumerate the six human traits which you most frequently use in rating your fellow students.

After question 1 was given and the answers were written on the paper provided, the second question was dictated. The students were told to use all the time they wanted, but to write down the traits as rapidly as they thought of them. Each member of the class was requested to prepare his list without communication with other members. This request was diligently observed. At the end of twelve minutes all students had completed their lists, and were very anxious to report and hear others report. Each student desired to know whether or not his list if six human traits had been partially or wholly reproduced by some other member or members of the class.

Student No. 1 was called upon to place on the board before the class the list of six human traits most frequently used by him in judging the intelligence of his fellow students.

Student No. 2 was asked to add to the list on the board any trait or traits from his list of six which were not mentioned by Student No. 1. Thus we proceeded through the entire class. Should each student use a different list of human traits in forming his judgments, and thereby add six traits to the list on the board, 180 human traits would be enumerated on our master list. But this did not happen. No one could expect this to happen. However, we doubt if many would think that students use so nearly the same criteria in judgment that, of the 180 possibilities only 13 were actually enumerated by the entire class. And under discussion of the class this list of 13 traits was reduced to 6. The 13 traits listed below are given in the same order as they were listed on the board by the students.

1. Quick perception.
2. Adaptability.
3. Leadership in student group.
4. Ability to reason beyond self-interest.
5. Mental alertness.
6. Retention of learning—memory.
7. General appearance and manner.
8. Social poise.
9. Laughter.
10. School grades.
11. Philosophy of life.
12. Originality in thinking.
13. Enjoyment of play.

In the opinion of the class there was overlapping in these characterizations. Therefore, through considerable discussion and definitions we were able, as stated above, to reduce this list to six. These six traits which were to form the students' master scale are given below:

1. Leadership among students.
2. Impersonal reasoning[2]—ability to reason beyond one's own selfish ends.
3. Mental alertness—quickness and ease with which novel material is grasped.
4. Appearance and manner—ability to present one's self to win confidence and respect.
5. School grades: Grades on all college courses.
6. Originality in thinking: Habit of thinking and doing things in an unusual way.

A careful study of the six remaining traits reveals the fact that they, when properly defined, include the 13 former characterizations of the original list: No. 1 is included under No. 3 of the final list; No. 2, under 1 and 3; No. 6, under 5; and Nos. 8, 9, 11, and 13 under 1, 3, and 4. The class was in full agreement in every instance, except in the case of No. 9—laughter. All agreed that "laughter" could be placed under "appearance and manner," yet three students insisted that as a trait for diagnosis of intelligence it is too important to be subordinated in any way.

THE STUDENTS' RANKING OF THE SIX TRAITS SELECTED.

At another meeting of this same class the students were requested to rank these six human traits on the basis of their importance in determining intelligence. Each student was provided with a printed list of the six traits, and asked to follow directions carefully as they were read. "Do exactly what I tell you. Select that one of the six human traits listed here which you think is of greatest value in judging intelligence in a fellow student. Mark this trait 'highest.' Now, select that one of these six traits which you think is of least value in judging the intelligence of your fellow students. Mark this 'lowest.' With this done, select that trait midway between highest and lowest in value and mark it 'middle.' Select that trait now which is of less value than the one marked 'highest' but of greater value than the one marked 'middle.' Mark this trait 'high.' Then select a trait which is not so good as middle but better than lowest. Mark this one 'low.' Take the remaining trait of the original six and compare its value with that of each of the other five. Place it between those two which it is most like."

---

[2] During the class discussion it was made plain by the students that by "impersonal reasoning" they meant the ability to solve a problem to its conclusion without clouding the issue by the intrusion of the personal element.

When this ranking had been completed by the class and the results tabulated, the final six traits were arranged in the order of their relative importance in determining the intelligence of a fellow student. Their relative positions are:

1. Mental alertness.
2. Appearance and manner.
3. Originality.
4. Leadership.
5. Impersonal reasoning.
6. School grades.

By accepting these six traits as the criteria for judging intelligence we were able to use this as a master scale for ranking college students.

Some objection may be offered to the acceptance of these traits as criteria for measuring intelligence. It may be urged against some one, perhaps any one, of these traits that there is *no relationship between it and intelligence. Question: Are these* traits earmarks of intelligence? To what extent these several traits are characteristic of intelligence it is difficult to say. Yet, the following two types of evidence persuaded us to accept them as worthy criteria. First, they represent the consensus of opinion of an average group of college students as to what they think intelligence is. That they had the correct conception of the term "intelligence" we are not arguing. We do think, however, that they had in mind what they call "intelligence," and what they constantly use in placing their fellow students intellectually, and their success in this placement is the aim of this study. The second type of evidence is found in the correlation between these several traits and the results of the Otis Test. When the students were rated on these traits, as is shown further on in this study, and compared with their rank in terms of the Otis Tests the respective value of these traits as characteristics of intelligence is exhibited. The Otis Tests correlate with "mental alertness" .49, with "appearance and manner" .16, with "originality" .48, with "leadership" .24, and with "impersonal reasoning" .38. While "appearance and manner" was considered by the students as the second best characteristic of intelligence, this trait does not seem to be important to success in the Otis Test. There is a reason for this. And it it this reason *which causes many people to discountenance intelligence tests* and hark back to the practice of the personal interview. And for the same reason others hold that the intelligence test and the personal interview are mutually supplementary.

CLASS MEMBERS RATED BY MASTER SCALE.

a. *By the Students.* The class roll in strictly alphabetical order was printed on the left side of a large sheet of paper. Near the top, the middle, and the bottom of the page was printed 'highest,' 'middle,' and 'lowest,' respectively. Midway between 'highest' and 'middle' the word 'high' was printed, and midway between 'lowest' and 'middle' the word 'low' was printed. One such sheet was provided for each of the six traits of the master scale. The class was directed to rate its members in the same way that it rated the six traits of the master scale on the day before. When the class was rated on *mental alertness*, the papers were gathered up. Two days later data were secured in the same way on appearance and manner. This two-day interval was allowed that we might avoid the chance of habitually assigning a man to a certain place on the scale. Only five of the traits were rated by the students, since the college grades were secured from the university records.

By an examination of Table No. 1 several interesting points are observed. 1. In the judgment of the class no one student possessed all five traits in the highest degree. Student No. 11, who makes first place in three traits, is the nearest an exception to this. 2. One may rank high in one trait and low in another. Student No. 8, for example, is given sixth place in mental alertness and twenty-sixth in appearance and manner, and No. 3 ranks second in impersonal reasoning and twenty-seventh in appearance and manner. Students 25, 26, and 27 also indicate the truthfulness of this statement.

This observation is not in line with the "halo" theory as presented by Thorndike. His findings seemed to indicate that those giving the ratings were unable to analyze out these different aspects of a person's nature and rate each independent of the others. "All ratings were effected by a marked tendency to think of the person in general as rather good or rather inferior and to color the judgments of the qualities by this general feeling. A halo of general merit is extended to influence the rating for the special ability, or vice versa."[3] While the data here do not indicate the presence of this tendency generally, yet in some outstanding cases it seems to have been operative.

---

[3] Thorndike, E. L.: *A Constant Error in Psychological Ratings*, Jour. Appl. Psychl., March, 1920.

TABLE NO. I.

STUDENTS' ESTIMATES OF INTELLIGENCE

Student No.	Mental Alertness	Appearance and Manner	Originality	Leadership	Impersonal Reasoning	Rank
1	23	28	21	30	15	28
2	25	7	14	11	24	15
3	16	27	22	26	2	20
4	18	9	19	13	9	10
5	22	17	30	17	20	25
6	9	13	15	21	6	9
7	19	6	13	8	11	8
8	6	26	16	16	14	12
9	20	5	20	12	16	11
10	12	3	7	6	5	5
11	1	4	1	5	1	1
12	11	1	3	2	7	3
13	14	22	8	19	17	13
14	27	19	10	7	21	16
15	21	20	25	29	19	27
16	10	14	2	1	3	4
17	15	23	26	23	10	21
18	24	10	24	18	30	26
19	8	15	11	10	12	7
20	29	21	29	24	28	30
21	2	12	28	20	25	19
22	28	11	27	26	29	29
23	17	25	23	9	23	22
24	13	30	5	27	26	23
25	3	24	17	14	22	14
26	5	29	16	28	8	17
27	4	16	13	3	13	6
28	7	2	4	4	4	2
29	26	18	9	15	18	18
30	30	8	18	22	27	24

To show more fully the extent to which the several traits *correlate with each other the table of intercorrelations is given* below.

TABLE NO. II.

INTERCORRELATIONS OF TRAITS.

	Mental Alertness	Appearance and Manner	Originality	Leadership	Impersonal Reasoning
Mental Alertness	......	.03	.40	.31	.55
Appearance and Manner	.13	......	.35	.54	.20
Originality	.41	.34	......	.60	.53
Leadership	.31	.54	.60	......	.39
Impersonal Reasoning	.55	.20	.53	.39	......
Average	.32	.28	.47	.46	.42

With this group of students "appearance and manner" is least indicative of mental alertness. In fact, the two traits do not appear to hold together. From Table No. 2 we also note that "originality" shows the highest while "appearance and manner" shows the lowest intercorrelation with these specific traits. Then "originality" would therefore, give us the most information, and "appearance and manner" the least information about this whole group of traits.

b. By the Instructor. Using the same method as the students, the instructor rated the members of the class on each trait of the scale the results are given in Table No. III.

TABLE NO. III.

INSTRUCTOR'S ESTIMATES OF STUDENTS.

Student No.	Mental Alertness	Appearance and Manner	Originality	Leadership	Impersonal Reasoning	Rank
1	27	7	28	21	27	24
2	13	16	17	16	14	14
3	7	17	21	11	9	11
4	29	24	23	20	19	27
5	8	9	15	14	13	9
6	12	4	10	13	12	7
7	23	28	29	25	25	30
8	1	1	7	2	1	1
9	21	18	24	17	20	22
10	9	11	3	12	10	6
11	3	6	2	6	11	3
12	10	3	12	3	5	4
13	16	10	14	5	17	10
14	14	22	20	8	21	16
15	24	12	25	28	22	25
16	6	5	1	1	6	2
17	17	19	16	15	28	20
18	28	8	30	26	29	28
19	4	14	5	10	7	5
20	15	20	26	18	8	17
21	20	13	22	23	18	21
22	22	15	27	24	26	26
23	25	25	19	4	15	18
24	18	23	4	27	30	23
25	5	26	9	22	4	12
26	11	30	18	29	3	19
27	2	21	8	19	2	8
28	19	27	13	7	16	15
29	30	29	11	30	23	29
30	26	2	6	9	24	13

Observations made from Table No. I also apply here. When the rating of a student in one table is compared with that student's rating in the other, this observation is made. Only six students are rated the same on any one trait in both tables. Of these, four are on leadership. They are Nos. 11, 14, 17, and 25. This agreement, though slight as it seems, selects the highest ranked leader and almost the lowest.

The intercorrelations of the traits as presented by the instructor's estimates are shown in Table No. IV.

TABLE NO. IV.

INTERCORRELATIONS OF TRAITS.

	Mental Alertness	Appearance and Manner	Originality	Leadership	Impersonal Reasoning
Mental Alertness	......	.29	.85	.49	.80
Appearance and Manner	.29	......	.25	.48	.14
Originality	.85	.25	......	.70	.45
Leadership	.49	.48	.70	......	.43
Impersonal Reasoning	.80	.14	.45	.43	......
Total	.61	.29	.56	.51	.46

"Mental Alertness" shows the highest intercorrelation of all the traits with "originality," the students' most significant trait, taking second place. "Appearance and manner" continues to be the lease significant of all the traits.

CORRELATION OF STUDENTS' ESTIMATES OF INTELLIGENCE WITH STANDARD SCALES.

In order to test the value of students' estimates of the intelligence of fellow students by means of these traits, we have brought together in Table V the results of Tables I and III and the students' ranking in both college grades and the Otis Intelligence Test.

In column No. 1 the students are ranked by the class on the basis of the six traits in Table I. Column No. 2 contains the ranking of the same students by the instructor on the basis of these same six traits. In column No. 3 the students are ranked according to their grades in the office of the University. The ranking of the students on the Otis Group Intelligence tests is presented in column No. 4. Their scores were made on the "Advanced Examination: Form A. Edition 1921."

The instructor's estimate and college grades are combined for the ranking given in column No. 5. This is done to get a comparison of this combined rating with the results on the Otis Test. Before attempting to compare the estimate made by the students to any of these standards, I think it wise first to determine the value of the instructor's estimate, the college grades, and the Otis Intelligence tests in relation to each other. Since the Otis test is generally accepted as a standard measure of intelligence, I will evaluate the others in terms of it. By comparing column 2 with column 4, one secures a correlation of .46, between the instructor's estimate and the Otis Test. In the same way the relation between college grades and the results on the Otis test is found by comparing columns 3 and 4. This correlation is .53. By a comparison of the rankings in columns 2 and 3 the correlation between instructor's estimate and college grades is found to be .41. The instructor's estimate and college grades are often combined in judging intelligence. Their combined value when compared with the Otis test is found by comparing columns 4 and 5. The correlation is .57. When combined in this way they give a higher correlation with the Otis test than either one when taken alone. These correlations are "markedly present," yet not too high. However, since both instructor's estimate and college grades are in such common use I am using them as standards along with the Otis Intelligence Test.

These three standards are significant because of the fact that they are of different types. The Otis Test is a theoretical attempt to build a scale of measurement. Its further value lies in the fact that its scores check favorably with scores made on other tests which are equally sound in theory. The instructor's estimate is an entirely different kind of standard. It is an individual judgment, while the students' estimates form a composite judgment. Theoretically this last should be the proper basis for judging, for after all this is the final criterion in the business world. It is the composite judgment of one's fellows that places one at a particular level. Hollingworth finds that group judgment is more accurate than the judgment of the individual. "The individual judges intelligence with a correctness of only .19 on the average; the group judges intelligence with an average correctness of .51." [4]

a. Instructor's estimate. By a comparison of columns 1 and 2 of Table V one is able to determine the relationship between the students' estimate of intelligence and that of the instructor. This relation, expressed by means of correlation is .61. This is a rather high correlation and would seem to justify the conclusion that the summation of the opinion of students with the maturity and college rank of this class is as good a measure of intelligence as their instructor's judgment.

TABLE NO. V.

ESTIMATES AND STANDARD SCALES.

Student Number	Students' Estimates	Instructor's Estimates	College Grades	Otis Tests	Inst. Estimate and College Grades
1	28	24	22	20	26
2	15	14	30	14	25
3	20	11	8	10	7
4	10	28	9	4	17
5	25	9	7	27	5
6	9	7	16	12	11
7	8	30	27	24	30
8	12	1	3	1	1
9	11	22	25	29	27
10	5	6	15	17	9
11	1	3	1	2	2
12	3	4	5	16	3
13	13	10	17	23	13
14	16	16	26	18	23
15	27	25	28	30	29
16	4	2	29	6	14
17	21	20	18	21	20
18	26	28	13	22	21
19	7	5	4	3	4
20	30	17	20	28	18
21	19	21	21	19	22
22	29	26	24	26	28
23	22	18	19	15	19
24	23	23	10	8	15
25	14	12	6	13	6
26	17	19	2	7	10
27	6	8	11	5	8
28	2	15	12	25	12
29	18	29	14	9	24
30	24	13	23	11	16

b. College grades. When the rating of the members of the class made by the students and found in column 1 is compared with the results based on college grades and given in column 3, the correlation is found to be .34. This is in keeping with the student judgment, for when building their scale they placed college grades as the poorest of the six traits for measuring intelligence. One must not confuse the student's ability to make college grades with what he actually does. These students are ranked on the grades

---

[4] Hollingsworth; *Judging Human Character'* page 41. Similar findings are reported in other chapters of this book.

they have actually made. We cannot say how nearly they approach the best of their several abilities.

c. Otis Intelligence Tests. As a final measure of the students' estimates of intelligence the students are ranked according to their success in the Otis Intelligence Tests. The range of their scores on this test was from 125 to 203. The correlation of the two rankings thus given in columns 1 and 4 in Table V is .51. Students estimates of intelligence compare very favorably with the results of the Otis Test, slightly better, in fact, than other investigators have found college grades to relate to intelligence. Pressey obtained a correlation of .48 between intelligence and school marks.[5] Terman found a correlation of .45 between school marks and mental age.[6] Book reports the coefficient of correlation for high school seniors between school marks and intelligence to be .47.[7] If, therefore, we take the Otis Intelligence Test as a standard the consensus of student opinion must be highly respected as a measure of intelligence.

SUMMARY.

1. Correlation Summary:

	Students' Estimates	Instructors' Estimates	College Grades	Otis Tests	Col. Grades & Inst. Estimates
Students' Estimates	.....	.61	.34	.51	.48
Instructor's Estimates	.61	.....	.41	.46	.82
College Grades	.34	.41	.....	.53	.83
Otis Tests	.51	.46	.53	.....	.57
Col. Grades and Inst. Est.	.48	.82	.83	.57	......
Average	.49	.59	.52	.52	.67

2. If the instructor's estimate is an accepted standard, the consensus of student opinion, gained under the conditions of our experiment, is a better measure of intelligence than college grades or the Otis Tests.

3. If the college grades are taken as the accepted standard the students' estimates are not so good a measure of intelligence as either the instructors' estimates or the Otis Tests.

4. Accepting the Otis Tests as a standard, the students' estimates are practically as good a measure as either the instrustors' estimates or the college grades. If, however, the instructors' estimates and college grades are combined, they give a slightly better measure.

---

[5] Pressey, S. L.: *The Efficiency of the Group Point Scale in Prognosticating Success and Failures in Junior High School.* Jour. of Applied Psychology, 3:381-385.

[6] Terman, L. M.: *The Intelligence of School Children*, 1919, page 79.

[7] Book, W. F.: *The Intelligence of High School Seniors*, 1922, page 105.

A SCHEME FOR COMBINING INCOMPLETE RANKINGS

BY B. F. HAUGHT
The State University of New Mexico

School marks are usually assigned to several groups of students in terms of letters. The number of groups ranges from three to seven. When these marks are to be combined in such a way to get the student's average standing, the task is found perplexing and often ambiguous. Sometimes percentage values are assigned to each letter and these values averaged. In other cases point valuse as 5, 4, 3, 2, 1 and 0 are assigned to the letters A, B, C, D, E, and F respectively and the average found. Both of these methods often give very unsatisfactory results. The principal cause for this is the variation in the absolute marks assigned by different instructors. In some classes where there is no reason to infer that the selection has been great, a large per cent of the class receives high grades and at the same time in other classes the low grades are more numerous. In order that the work for the semester or term may be averaged in a satisfactory manner, it is desirable to find some scheme that will partially eliminate this wide variation in teachers' judgments.

Teachers agree much more closely in ranking a group of students or a bunch of examination papers than they do in assigning absolute grades in terms of letters or percentages. Usually there is no difficulty in ranking the best and the poorest fourth of the class. It is not usually difficult to rank a class of thirty with considerable accuracy. The smaller the class the easier the task unless the process of selection has been very great. Even then, if vigorous work is required and the tests are made difficult enough, individual differences are revealed.

That instructors do vary greatly in their marks does not need further proof to those who have had experience with teachers' grades. That rankings made by different teachers are less variable and consequently much more reliable may not be accepted so readily. Meyer published the distribution of marks assigned by forty professors at the University of Missouri to their students in a period of five years. These distributions show that the per cent of A's assigned ranges from 1 to 55. Similar tabulations have been published for Harvard University, University of Wisconsin, and Cornell University [1]. All these tabulations furnish evidence of the wide variation in marking. Starch presents further evidence to show that the variation among teachers in marking the same paper is just as great. When school work in two subjects like arithmetic and language, however, is correlated, a marked or high correlation is usually obtained [2]. This is evidence that, although the teachers have varied greatly in the specific marks assigned, they have not varied so much in the order of ranking.

---

[1] Starch, Daniel: *Educational Psychology;* 1920, 420-31.

If we admit that ranks are more accurate than the traditional school marks, are we any nearer the solution of our problem? How are we to combine these ranks? No instructor will rank all the students in any one class. How shall we weight a rank of 1 in a class of 5 and the same rank in a class of 35? Certainly they should not have equal weight.

The first investigation that attempted to solve this problem, as far as the writer knows, was by Professor E. L. Thorndike [3]. Recently other investigations have appeared. Important among these is the one by M. J. Ream [4]. Ream's method is only a modification of the one set forth earlier by Thorndike, but certainly is better in being more accurate and less laborious. The scheme here presented is only a modification of that suggested by Ream. It is better for two reasons. First, it eliminates negative signs and decimal fractions, both of which are likely to cause errors, to say nothing of the extra time and care. Second, the units are expressed in percentiles instead of in standard deviations. The former has more meaning to the average person. It is, of course, true that they amount to the same thing.

It is assumed that all distributions are normal and that any skewness is due to the teacher's inaccuracy in grading or marking. This assumption is not absolutely true, but it is much truer than to assume the distributions obtained from the teachers' grades correct. In the writer's opinion, to throw a distribution into percentiles and thus give it the form of the normal frequency curve corrects many errors of judgment on the part of the teacher. It will have to be admitted, however, that it introduces some errors as well as retains some of those already there. This is especially true when the group being considered is highly selected.

A concrete example of how the percentiles are found for a class of six will make the procedure clear. First, the numbers 1, 2, and 3 are divided by 6, the number in the class. This gives 17,

---

[2] Starch, Daniel: *Educational Psychology*; 1920, 51-57

[3] Thorndike, E. L.: *The Technique of Combining Incomplete Judgments of the Relative Position of n Facts Made by n Judges.* Journal of Philosophy, Psychology and Scientific Methods, 1916, 13, 197-204.

[4] Ream, M. J.: *A Statistical Method for Incomplete Order of Merit Ratings.* The Journal of Applied Psychology, 1921, 5, 201-08.

33 and 50. Second, from Thorndike's Mental and Social Measurements, Table 54, the numbers corresponding to 17, 33 and 50 are 1.49, .68 and .22 respectively. Third, 3 is added to each of these numbers to give the distance in terms of the standard deviation from the lower end of the distribution instead of from the middle. Fourth, each of the three sums is divided by 6, the length of the distribution in terms of the standard deviation. The quotients, 75, 61 and 54 are the percentiles. The scores for the lower half of the group may be found by inspection since the distribution is assumed to be symmetrical.

Method of application. Now suppose we wish to find the average score of an individual. Let us take an actual case, that of student X. He ranks

1 in a class of 6
1 in a class of 42
2 in a class of 18
2 in a class of 22
13 in a class of 40

By consulting the tables, we find the corresponding percentiles to be 75, 91, 73, 75, and 58 respectively. The average of these is 74.4. In a like manner the average may be found for all other students.

At the University of New Mexico, each instructor is requested to rank the upper half of his classes and include these rankings in his final semester report. For the semester ending January 28, 1922, the writer asked each instructor to rank the lower as well as the upper half of the class. Two exceptions were made. The instructor of history was not able to have charge of his classes the last month. They were taught by substitutes and for this reason are not included. The other exception was music, which probably depends more upon special ability than upon general intelligence.

After all the data were assembled, the average for each student was found in a manner similar to that used in the illustration above. All subjects that were not ranked as many as three times were discarded. No attempt was made to separate the college classes. If freshmen and seniors happened to be in the same class, they were ranked without respect to classification. No distinction was made for classes meeting two, three, four, or five times a week. This could have been done, but seems foreign to the problem.

During the semester the writer gave the Alpha tests to 148 students in the University. A few of these dropped out before the end of the semester. Altogether there were 138 who had scores in scholastic achievement and in the Alpha tests. The 138 students were assigned percentile ranks in each trait. These were assigned in such a way as to make the distribution take the form of the normal probability curve.

As a check on this method of finding scholastic scores, the following values were assigned to the several letters and the student's average grade found:

A=94%
B=88%
C=81%
D=74%
E=67%
F=50%

After these averages were found, the 138 students were assigned percentiles so as to give the distribution the form of the normal probability curve. Three correlations (by the product-moment method) were now obtained: first, between intelligence tests and the scholastic standing determined wholly from the ranks; second, between the intelligence tests and the scholastic standing determined by assiging absolute values to the letters indicating the grades; and third, between the scholastic standings determined by the two methods indicated above. The three correlations are repectively, as follows;

$$r_1 = .40$$
$$r_2 = .35$$
$$r_3 = .78$$

The method of disregarding absolute grades and using only the ranks, then, gives a more reliable result judged from the intelligence tests as a criterion. The difference between the two methods is small, yet it is sufficient to be significant when this number of subjects is used. The two scholastic rankings have a high correlation, yet it is far from perfect.

Great scientific accuracy is not claimed for the scheme here set forth. It is as accurate as any now in use and has the advantage of being very simple and easy to manipulate. The smaller the class the less reliable are the results. In classes as small as four or five no great injustice seems to be done to any one unless all the members are very good or very poor, which is not likely to happen.

TABLE I.

	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	68	71	73	75	77	78	79	79	80	81	81	82	82	83	83	85	85	85	85	86	86	86	86	86	86	86	88	88	88	88	88	88	89	89	89	89	89	89	91	91	91	91	91	91	91	91	91	91
2	50	55	59	61	64	65	66	67	68	69	70	71	72	73	73	73	74	74	75	75	75	76	76	76	76	76	77	77	77	77	77	77	78	78	78	78	78	78	82	82	82	82	82	82	82	82	82	82
3	32	45	50	54	56	58	60	61	62	63	65	66	66	67	67	68	69	69	70	70	71	71	71	71	72	72	72	73	74	74	74	74	74	74	75	75	75	75	76	76	76	76	76	77	77	77	77	77
4	—	39	41	46	50	53	55	57	58	59	61	62	62	63	64	64	65	66	66	67	67	68	68	68	69	69	70	70	70	71	71	71	71	71	72	72	72	72	74	74	74	74	74	75	75	75	75	75
5	—	—	27	39	44	47	50	52	54	55	57	58	59	60	61	61	62	63	63	64	64	65	65	66	66	67	67	68	68	68	68	69	69	69	69	69	70	70	70	70	71	71	71	72	72	72	72	72
6	—	—	—	23	30	42	45	48	50	52	53	55	56	57	58	59	59	60	61	61	62	62	63	64	64	64	65	65	66	66	66	66	67	67	67	68	68	68	69	69	69	70	70	70	70	70	70	70
7	—	—	—	—	23	35	40	43	46	48	50	52	53	54	55	56	57	58	58	59	60	60	61	61	62	62	63	63	64	64	64	64	65	65	66	66	66	66	67	67	67	67	68	68	68	69	69	69
8	—	—	—	—	—	27	34	39	42	45	47	49	50	51	53	54	55	55	56	57	58	58	59	59	60	60	61	62	62	62	63	63	63	64	64	64	64	64	65	65	66	66	66	67	67	67	67	67
9	—	—	—	—	—	—	21	32	37	41	43	45	47	49	50	51	52	53	54	55	56	57	57	57	58	59	59	60	60	61	61	62	62	62	62	63	63	63	64	64	64	65	65	65	65	65	66	66
10	—	—	—	—	—	—	—	21	31	36	39	42	44	46	47	49	50	51	52	53	54	55	55	56	57	57	58	58	59	59	60	60	60	61	61	61	62	62	62	63	63	64	64	64	64	64	65	65
11	—	—	—	—	—	—	—	—	20	31	35	39	41	43	45	46	48	49	50	51	52	53	53	54	55	55	56	57	57	58	58	59	59	59	60	60	60	61	61	61	62	62	62	63	63	63	64	64
12	—	—	—	—	—	—	—	—	—	19	30	34	38	40	42	44	45	47	48	49	50	51	52	52	53	54	55	55	56	56	57	57	58	58	59	59	59	59	60	60	60	61	61	61	62	62	62	62
13	—	—	—	—	—	—	—	—	—	—	19	29	34	37	39	41	43	45	46	47	48	49	50	51	52	52	53	54	55	55	55	56	56	57	57	58	58	58	59	59	59	60	60	60	60	61	61	61
14	—	—	—	—	—	—	—	—	—	—	—	18	27	33	36	39	42	42	44	45	46	47	48	49	50	51	52	52	53	54	54	55	55	55	56	57	57	57	57	58	58	58	59	59	59	60	60	60
15	—	—	—	—	—	—	—	—	—	—	—	—	18	27	33	36	38	40	42	43	44	45	47	47	48	49	50	51	52	52	53	53	54	54	55	55	56	56	57	57	57	57	58	58	59	59	59	59
16	—	—	—	—	—	—	—	—	—	—	—	—	—	17	27	32	35	37	39	41	42	43	45	46	47	48	48	49	50	52	52	52	52	53	53	54	54	55	55	56	56	57	57	57	57	58	58	59
17	—	—	—	—	—	—	—	—	—	—	—	—	—	—	17	27	31	34	37	39	40	42	43	44	45	46	47	49	49	49	50	51	52	52	52	53	53	54	54	55	55	56	56	56	57	57	57	57
18	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	15	26	31	34	36	39	40	41	42	43	45	45	46	47	48	49	49	50	51	51	52	52	53	53	54	54	54	55	55	56	56	56	57
19	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	15	26	30	33	36	38	39	41	42	43	44	45	45	46	47	49	49	49	50	51	51	52	52	53	53	53	54	54	55	55	55	56
20	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	15	25	30	33	35	37	39	40	41	42	43	44	45	46	47	48	49	49	49	50	51	51	52	52	52	53	53	54	54	54	55
21	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	15	25	29	32	35	36	38	40	41	42	43	44	45	45	46	47	48	48	49	49	50	51	51	52	52	53	53	53	53	54
22	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	14	25	29	32	34	36	38	39	40	41	42	43	44	45	46	47	47	48	48	49	49	50	51	51	52	52	52	53	53
23	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	14	24	29	32	33	36	37	38	40	41	42	43	44	45	45	46	47	47	48	49	49	49	50	51	51	51	52	52
24	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	14	24	29	31	33	35	37	38	39	40	42	42	43	44	45	46	46	47	47	48	48	49	49	50	51	51	51
25	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	14	24	29	32	33	35	36	38	39	40	41	42	43	43	45	45	46	46	47	47	48	48	49	49	50	51
26	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	14	24	28	30	31	34	36	37	39	40	41	42	42	43	44	45	45	46	47	47	47	49	49	49	49
27	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	14	21	26	30	32	34	36	37	39	39	40	41	42	43	43	44	45	46	46	47	47	48	48	49
28	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	14	22	27	30	32	34	36	37	38	39	40	42	42	43	43	44	44	45	46	46	47	47	48
29	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12	22	25	29	32	34	35	36	38	39	40	41	41	42	43	43	44	45	45	46	47	47
30	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12	22	26	29	32	33	35	36	37	38	39	40	41	42	43	43	44	44	45	46	46
31	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12	22	26	29	31	33	34	36	37	38	39	40	41	42	42	43	43	44	45	45
32	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12	22	26	29	31	33	34	36	37	38	39	40	40	41	42	43	43	44	44
33	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12	21	26	28	31	31	34	36	36	37	38	39	40	41	42	42	43	43
34	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12	21	26	29	31	32	34	35	36	37	38	39	40	41	41	42	43
35	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12	22	25	28	30	32	33	35	36	38	39	39	40	40	41	42
36	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12	22	25	28	30	31	33	34	35	36	37	38	39	40	41
37	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12	22	25	29	30	31	33	34	35	36	37	38	39	40
38	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12	22	25	28	30	31	33	34	35	36	37	38	39
39	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12	22	24	26	29	30	32	33	35	36	36	38
40	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	12	18	24	26	29	30	32	33	35	35	36
41	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9	18	22	26	29	30	32	33	34	35
42	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9	15	24	26	29	30	31	33	34
43	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9	18	22	25	28	30	31	32
44	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9	18	22	25	28	30	31
45	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9	15	22	25	28	30
46	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9	19	22	25	26
47	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9	18	23	25
48	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9	15	22
49	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9	18
50	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	9

Showing percentile scores as determined by the number in the class and the rank in that class. Numbers at the top of each column indicate the number in the class and numbers in the column at the left indicate the ranks.

THE BASIS OF TOOTH PASTE SALES IN REPRESENTATIVE COMMUNITIES

BY DONALD A. LAIRD.

I. INTRODUCTION

This is a report of a fairly extensive survey which was made of two distinct communities with reference to the use and the sales of tooth pastes and similar preparations. The survey was made by a class in the psychology of advertising as a term project. It was conducted in Laramie, Wyoming, a town of about seven thousand inhabitants. The population is composed in the main of retired ranchers, business and professional families, and the families of workers in the refinery, plaster mills, Union Pacific division employees, and a scattering of students.

It was not attempted to cover the entire town. Rather the aim was to select as rigorously as possible certain sections which represented unequivocally sociological extremes typical of the Main Street town. One section of the town was chosen for study because it represented the *elite*, contained the residences of the better classes, business and professional men, university families, and a scattering of well to do ranchers who manage their land affairs from the town. This section was purposely kept small in order to be certain that the families who contributed information for this group were without doubt properly labelled. In a small town there is little difficulty in so classifying the townspeople.

Opposed to this section, the *creme de la creme* of Laramie, a section was mapped out which contained the colored section of the population, a majority of the foreign element, and the unskilled workers in general. This section is commonly referred to as the *West Side*. In this report the two groups will be designated the Elite and the West Side.

II. THE SURVEY.

A house to house canvas was made of the two carefully selected districts. For each house information was gathered under the following beads:

1. Street and number
2. Persons in family (excluding roomers)
3. Brand of paste used
4. Why used (flavor, dentist's advice, samples, advertising, friends, etc.)
5. How many tubes used a month
6. How many in family use paste once daily

7. Any other brands used in family
8. Why others are used

The canvasers were instructed to pass by any homes which seemed by outward appearances not to properly belong in the district being surveyed.

The students were divided into two equal groups, one for each section. A student was selected from out of each group to serve as section captain in taking charge of the details of the survey.[1] Several students not in the class requested an opportunity to take part in the data gathering. The only canvaser to be thrown out of a house was one of these volunteer solicitors who was ejected bodily, even though the door-mat read "Welcome!"

III. RESULTS.[2]

A total of 228 families were interviewed, 104 of which were from the Elite district. This report is based, however, upon only 148 families. Some were eliminated due to grossly incomplete data, some others were also eliminated from the Elite returns because of bad ratings in the local credit bureau. Eighty-one of the families to be reported were from the West Side.

Table I gives the distribution of the families arranged by number in the family.

TABLE I.

DISTRIBUTION OF SIZE OF FAMILY IN THE TWO DISTRICTS

	Frequency	
Number in family	Elite	West Side
1	0	1
2	18	19
3	18	16
4	12	14
5	14	14
6	1	9
7	2	3
8	1	3
9	1	0
10	0	1
11	0	1

The medium size with the Elite group falls in the 3 interval, with the other group the medium falls in the 4 interval. The arithmetical mean for the Elite family size is 3.6, on the West Side it is 4.1.

In 67 families in the Elite section there are 54 that make use of a toothpaste, at least one member of the family uses it. In the other section there are 53 families out of the 81 in which at least one member uses tooth paste. On a per cent basis this is 80.6 per cent of the Elite families and 65.4 of the West Side families in which tooth paste is used.

---

[1] Section captains were Mr. Robert Johnson and Mr. Harry Engstrom.

[2] Tabulated with the assistance of Mr. Ralph Holtsma and Mr. Willard Brokaw.

It was difficult for many of those interviewed to conscientiously state, in reply to item 5, how many tubes were used a month in the family. In the Elite section there were only 20 families in which this reply was recorded. In the other district 45 families replied. These families represent 77 individuals in the Elite district and 194 individuals from the West Side. The twenty families from the Elite group report the use of 37 tubes (full size) a month. This is an average consumption of 0.48 tube a month per individual. It must be borne in mind that this is per individual for whom we have data available. If this included all the families it is quite probable that there would be a zero after the decimal point. The West Side families reported the consumption of 84 1-3 tubes per month. On an individual basis this is 0.43 tube per individual per month among those who reported.

These results are somewhat in contrast. Among all the families interviewed we have found that a considerably larger per cent from the Elite district use a dentrifice. But when this comparison is reduced to a comparison of amount used by those that do use a dentrifice then the two groups practically coincide.

Of most interest from the point of view of advertising is *why* the users use the particular brand that they do. Admittedly in many instances the replies to item 4 were rather forced, for they had never given the matter much consideration. Rationalizations in forced explanations were undoubtedly very conspicuous. But at that the reasons attributed by the users for their use are a revelation. Especially when checked up with the chief overt advertising appeals that are in vogue in the marketing of these preparations.

Table II gives a detailed tabulation of the responses to item 4. It should perhaps be noted in passing that before the solicitors went out into the field they were gathered together and instructed regarding the data to be collected. They were especially warned against suggesting any response other than what might be the natural one. In the case of item 4 the suggestions in parentheses were arrived at after the group making the survey had tabulated their own reasons for using their particular brand of dentrifice. In case it was necessary to give suggestions to the person being being interviewed they were instructed to read the entire list of suggestions before letting the subject respond.

TABLE II.

REASONS GIVEN FOR THE USE OF DENTRIFICE, BY BRAND AND SECTION

	Elite Section								West Section							
Brand	Totals	No Reason	Flavor	Easy to Get	Advertising	Recommended by Friends	Advice of Dentist	Cheap	Totals	No Reason	Flavor	Advertising	Recommended by Friends	Advice of Dentist	Samples	Cheap
Colgates	19	7	8	1	1	1	1	0	28	9	12	3	0	0	2	2
Pepsodent	16	6	7	0	1	0	2	0	13	7	3	2	0	0	1	0
Pebeco	0	0	0	0	1	0	0	1	0	2	1	2	0	4	0	6
Kolynos	3	0	1	0	0	0	2	0	4	1	0	1	1	1	0	0
Chamberlain's	1	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0
Chlorax	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Dr. Grave's	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Listerine	1	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0
Soda or salt	3	0	0	0	0	0	3	0	1	1	0	0	0	0	0	0
Dr. Lyon's	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	1
Nyal	0	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0
Totals		13	20	1	4	2	13	1		21	17	7	1	1	3	3

A comparison of the totals columns shows the difference between the two communities in the matter of paste used. The preponderence of Pepsodent and Pebeco in the Elite section is worth mentioning in passing. There are other interesting but minor differences in the choice of a dentrifice between the two groups.

When the reasons given for the use of the particular brand are examined group differences also appear. The largest difference under this head is in the advice of the dentists. This appeared 13 times in the Elite section, and 1 time in the other section. Samples, which had been given the children in school, was given as the reason for the start of the use of a particular brand in three homes of the West Side, and not once in the other section.

The predominant reason for the use of personal brands was that of the flavor of the preparation. Neglecting the no reason responses this reason of the flavor being liked has a greater frequency than all of the others combined. The ratio of 47:29 between the use of Colgates and Pepsodent can largely be explained on the basis of flavor. Even in the Elite section where one would expect to find the consumers influenced by scientific appeals in advertising we still find flavor predominating over dental advice and advertising in the reasons given from this group.

Undoubtedly the manufacturers are well aware of the importance of flavor in a dentrifice. Their advertising, however, almost without exception is directed toward certain scientific sounding appeals for the use of their product. When one gets down to the reasons given for the use of any particular brand, the scientific reason is not mentioned a single time. Undoubtedly it is the advertising, based upon the opinions of high dental authorities, that has established the rigorous tooth paste habit of Mr. Average Citizen, whether he lives on the West Side or on Easy Street. Flavor on the other hand, seems to be the principle determining reason for the use of any particular brand.

IV. SUMMARY.

A house to house survey was made of the two distinct districts by the students in a class in the psychology of advertising. A good and poor residential district was covered. The tooth paste habits of the inhabitants was enquired into. It was found:

1. This is an excellent method of instilling the spirit of scientific psychological approach to the problems of advertising;

2. About 80 per cent of the families from the better section and 65 per cent from the poorer section make some regular use of a dentrifice;

3. On the basis of individual consumption, there is no significant difference between the sections among those *that do use* a dentrifice. In other words, when the tooth paste habit once hits, it hits with equal force in both districts;

4. But, the tooth paste habit does not occur with as great frequency in the poorer district as in the better class district;

5. Dentist's advice plays but a very minor role in the use of dentrifices in the poorer districts;

6. In both districts the predominating reason for the use of *any particular brand* is determined largely by the flavor;

7. School samples play no apparent part in the use of a paste in the better districts, they do play some part in the poorer district.

VOCATIONAL TESTS FOR AGRICULTURAL ENGINEERS

BY H. E. BURT AND F. W. IVES
Ohio State University

INTRODUCTION

The following is a contribution to the development of vocational norms which are essential for scientific vocational guidance and selection. It presents nothing particularly novel in the way of technique but gives data for a vocation that has not hitherto been studied in this manner.

It is obvious in the university as well as subsequently that some men who undertake agricultural engineering as a vocation lack the requisite ability. Hence the problem of vocational prediction is in point. Moreover this vocation requires more than mere intelligence. Observation and analysis makes it clear that it is to quite an extent a question of special ability. So an effort was made to devise special mental tests for predicting this ability and to standardize them on advanced students in agricultural engineering.

METHOD

The selection of the tests was based on some preliminary experiments the previous year as well as upon further analysis of the mental factors involved in the vocation. Typical of such factors are motor coordination involved in the use of surveying and drawing instruments, various aspects of attention involved in testing farm machinery and taking readings at the proper moment, spatial perception and imagination involved in planning farm buildings with reference to their contents, and observation and judgment involved in planning topographical layout or drains. While no mental test measures any such factors exclusively, an effort to approach such measurement is far better than random selection of test material. Some of the tests employed were modifications of those used by various investigators and some were devised especially for the occasion. Those which proved useful and were retained in the final norms will be described below.

The tests were given, with two exceptions, by means of mimeographed blanks. All the blanks were clipped together in a folder. The group method was used and each test had a time limit such that no subject would quite finish it. Each test comprised two installments,—on successive pages or on the same page separated by a space and the word "stop," according to the amount of material. At one sitting the first installments of all the tests were given in succession. The following day the second installments were given in the same order. It was thus possible to locate cases where the subject misunderstood or made a bad start and also (if desirable) to correct for attenuation. The order of the tests was designed to avoid undue fatigue by interspersing tests requiring close concentration and effort with those somewhat less exacting. There were 14 tests used but as several were scored in more than one manner, there were 18 different independent variables involved in the research.

The directions were given orally. The first page of the folder contained examples of various tests and after finishing one test the subject marked the place in the folder with the finger and turned back to the example page during the explanation of the next. The examples were in most cases designed to allow one or two items for practice. The directions were, of course, standard and given verbatim but were abbreviated somewhat for the second installments.

The subjects were 43 advanced students specializing in agricultural engineering at Ohio State University. They were tested in two groups during the winter of 1921-22. They were directed to take the examination by the dean of their college so it is safe to assume that incentive was sufficiently near a maximum to be constant. The testing was done personally by one of the writers (Burtt). The total time for the tests, including directions was a trifle over two hours, i. e. one hour at each sitting.

The criterion for evaluating the tests consisted of ratings as to "ability in agricultural engineering" made by three instructors in the department. Ratings were used rather than grades in engineering studies on the theory that the latter are less apt to reflect a man's actual ability.

The ratings were made on a linear scale of the sort commonly used. The names were typed in a column at the left of the page followed by five ruled columns of uniform width. These columns were headed from left to right "lowest fifth", "next lowest fifth" etc., and represented the distribution, on the basis of ability, of all the persons pursuing this vocation whom the judge new. A cross was to be placed after each name indicating the student's ability with reference to this whole distribution and the greater the ability the farther to the right the cross was to be placed. A preliminary statement at the top of the blank explained this method in detail. The ratings were then measured in millimeters from the left edge of the left column, thus expressing the judgment quantitatively.

The instructors made their ratings independently. The raw figures could not be validly averaged for a given subject because the judges might differ in the central tendency and scatter of their ratings. Consequently each judge's estimates were converted into the form *x* divided by *sigma*, i. e. the deviation of the measure in question from the mean of that judge's ratings taken in terms of the *standard deviation* of his ratings. With such converted measures the three judges' estimates for a given subject could then be averaged to get a combined rating (the criterion). Most of the subjects were rated by all three judges but in the case of three subjects only one rating was available.

The next step was the final selection of the tests. The scores for the two installments of each test were added together to form a single score which was used throughout. Correlation scatter plots were then made for each test against the criterion. It was found convenient to plot several pairs of variables simultaneously, keeping the criterion on the ordinate, aligning the blanks, placing a long ruler below the desired row and checking in the proper column on each blank for the other variables. Those plots which showed by inspection any semblance of correlation were computed by the "products-moments" formula. Eight of these were retained as showing some promise (the highest coefficient being .47) and intercorrelated and the regression equation derived.

DESCRIPTION OF TESTS

Only the seven tests which appeared in the final regression equation will be described. There were eight tests in the equation but the weighing of one of these proved to be zero. They are numbered as they were in the equation but these numbers have no ulterior significance.

Test 2 consisted of a series of small circles (letter "o" on a typewriter) in two columns 2½ inches apart. Those in each column were double spaced on the typewriter with the right column a single space below the left. They were numbered alternately in the two columns, i. e. odd numbers in the left and even numbers in the right. The subject connected the circles in the numbered order, i. e. drew straight lines from side to side, gradually moving down the page. He kept time with a metronome set at 120, i. e. made a stroke to the right and back again in a second. There were 56 circles in the column. The subjects were stopped immediately at the moment they should have finished so that if they did not keep with the metronome their blank was incomplete. There were two series of this sort on a blank. One was done as above described but for the second the page was turned 90 degrees so that the strokes were up and down. One such page constituted an installment of the test. The instructions were to stop as near as possible to the center of each circle but to keep time with the metronome. For scoring, a transparent stencil was made with a ¼ inch circle concentric with each "o". All strokes the end of which fell within this circle counted one point. This penalized those who did not follow the metronome to the extent of one point for each stroke they failed to make. The maximum score was 224.

Test 3 had at the top of the page a large diagram of a triangle, circle and rectangle overlapping in such a manner that there were areas comprising all three figures, each combination of two of them and each alone. In each of these areas were several symbols, —vowels, consonants and digits. The following are typical of items which appeared below the diagram:

1. Vowel that is in triangle but not in circle or rectangle ( )
2. Smallest number in circle and rectangle but not in triangle ( )
3. Consonant that is in rectangle circle and triangle ( )

There were 14 such items in each installment. The instructions were to place in the brackets at the end of each line the letter or number called for by the specifications in that line. The time limit was 2 minutes for each installment. The score was the number of items correct with a maximum score of 28 for the total test.

Test 4 was Thurstone's "Special Relations B" given in one installment only with a 4 minute time limit. The subject was given ample time to read the printed directions attached before turning to the test. The score was the number of items correct minus the number wrong. The maximum score was 39.

Test 5 comprised lines of digits in apparently random order,—65 digits to the line.[1] The test consisted in crossing out 2 and ringing 3 until coming to a 7 and then reversing the process,—crossing 3 and ringing 2,—until the next 7 etc. It was arranged, however, so that the first reaction after a 7 was always a "cross out." The subject knew this so if he became "lost" it was necessary to go back merely to the last 7. There were five crucial numbers between each pair of 7's and three such units per line. One unit comprised three 2's and two 3's and the next unit two 2's and three 3's etc. There were six lines in an installment and the time limit was 1 minute and 45 seconds. The subject was given one point for each correct mark. Furthermore in case several items in a unit were marked the opposite of what they should be and the subject had manifestly started the unit wrongly and reacted thereafter accordingly, he was given credit for the symbols thus marked but penalized one point for the initial mistake. The maximum total score was 180.

---

[1] A test similar to this has been used by H. A. Toops, but as far as the writers are aware has not been published.

Test 6 was auditory memory by the method of word pairs with logical material. There were three types of word pairs used; one type was normal, another comprised words related to agricultural engineering and a third involved pairs dealing with zeal, industry, interest in one's work etc. These three types rotated in succession through the list. The following is typical of a sequence of such pairs:

1	letter	stamp
2	formula	equation
3	task	eager
4	diamond	spade
5	liquid	hydrometer
6	strive	attain
7	spider	spin
8	engine	windmill
9	work	pleasure

There were 27 pairs in a series,—9 of each type. Two such series constituted an installment. The instructions were to listen while the 27 pairs were read and try to notice the natural connection between the two words. A metronome, somewhat muffled, sounded one beat a second as a guide for the experimenter. The two words read on two successive beats and there was a pause during the third beat. The first word of the next pair came on the fourth beat. The subject's blank consisted of a series of numbered lines. Immediately after the initial reading of the list by the experimenter, the first words of each pair were read with the appropriate number, e. g. "1 letter, 2 formula" etc. The subject wrote after the appropriate number the word that had paired with the one in question. Five seconds were allowed for each response. The score was the number of words correctly recalled. However the three types of pairs were evaluated separately. Especial interest attached to the total number of words of the agricultural engineering type recalled as compared with the total number of normal words recalled. The ratio of the former to the latter was the score used in the regression equation. The ratio was multiplied by 100 to simplify computation.

Test 8 was an "information" test comprising items related to agricultural engineering. Effort was made to select items that would not actually be discussed in the class room but with which an alert interested student who read current agricultural engineering literature would be familiar. Technical journals were used considerably in devising such items. The following are typical:

1. A track laying tractor is used for: ROAD ROLLING: PAINTING: MILKING: LAYING TRACKS: PLOUGHING.

2. The best blower belts are made of: LEATHER: RUBBER: JUTE: SILK: HEMP.

3. Soil stack is a term used in: BOILER FITTING: SURVEYING: CEMENT MANUFACTURE: PLUMBING: SOIL ANALYSIS.

There were 20 such items in each installment. The subject was instructed to underline the one of the five alternatives in capitals that belonged in the sentence. The time limit was 2 minutes per installment. The score was the number of items correctly marked and the maximum total score was thus 40.

Test 9 was a modification of Thurstone's directions test. Parts I and II had the following directions:

If the word contains E, A, and R mark it 1
If the word contains E but not A and R mark it 2
If the word contains A but not E and R mark it 3
If the word contains R but not E and A mark it 4
Parts III and IV had the following directions:
If the word contains I, N and G mark it 1
If the word contains I and N but not G mark it 2
If the word contains I and G but not N mark it 3
If the word contains N and G but not I mark it 4
If the word contains I but not N and G mark it 5

Each part consisted of 40 words of from 5 to 11 letters selected so as to belong to one of the classes indicated. They were arranged in 5 columns and each word was followed by a short dotted line for recording the answer. There were always an equal number of words requiring each sort of response in the "part". The subject was instructed to read the directions, look at the samples and mark the words in the same way. Parts I and III were done in successsion with time limits of 2 and 2½ minutes respectively and constituted the first installment. Parts II and IV were given likewise as the second installment. The score was the number correct and the maximum score 160.

RESULTS

The regression equation is as follows:

$$X_1 = .0021X_2 + .0195X_3 + .0091X_4 + .0133X_5 + .0128X_6 + .0070X_8 - .0014X_9 - 3.1892$$

Where X1 is the combined vocational score and X2—X9 the raw scores in the various tests above described. The weighting of *Test 7* proved to be zero. The coefficient of multiple correlation

$R_{1(23456789)}$ of the weighted sum of test scores and the criterion may be computed theoretically from the partial coefficients and proves to be .60. By actually weighting the original scores and correlating with the criterion the coefficient is .62±.07.

The extent to which the various tests contribute to the final weighted score may be seen from the following table which gives for each test the correlation with the criterion of zero order, e. g. r_{12}, and also of the seventh order with all the other tests constant, e. g. $r_{12.3456789}$.

Test	Zero order	Seventh order
2	.24	.10
3	.34	.11
4	.26	.13
5	.47	.35
6	.42	.30
7	.39	.00
8	.27	.05
9	.28	—.05

Tests 5 and 6 apparently carry most of the load. They have the highest coefficients of zero order and also the highest when the other tests are constant. *Test 7*, on the other hand, while being fairly good of itself merely duplicates factors involved in the other tests as shown by the partial coefficient of zero.

The correlation between *tests 5 and 6* is only .15. A regression equation was derived using only these two tests and is as follows: $X_1 = .0126X_5 + .0157X_6 - 2.64$.

The coefficient of multiple correlation $R_1(56)$ is theoretically .59 and the same result is obtained by weighting the original test scores. Moreover the weighted sum of *tests 5* and *6* correlates with the weighted sum of *tests* 2—9 to the extent of .92. So it is obvious that the two tests will give almost as reliable a prediction as the seven.

The technique of partial correlation involves, of course, the assumption of linearity of the original regressions. These regressions were not computed to test this assumption. However for the two final multiple correlations of criterion with weighted sum of tests the correlation ratio *(eta)* was computed. In the case of the weighted sum of the seven *eta* for x on y (y being the criterion) is .76, while r as above mentioned is .62. Blakeman's[2] criterion of linearity (using his simplest formula) is 2.1. Theoretically if this is less than 2.5 the regression may be considered linear. In the case of the weighted sum of the two tests *eta* is .74 and Blakemans's criterion is 2.2.

---

[2] Blakeman, J.: On Tests for Linearity of Regression in Frequency Distributions, *Biometrika*, 4, 332-350.

That the ability in question does not consist of mere intelligence but is definitely specialized is evident from the following. All the subjects had, on another occasion, taken the University Intelligence Test—a somewhat modified army alpha. Scores in this test correlate with the criterion to the extent of only $.16 \pm .10$, whereas the corresponding correlations for some of the special ability tests are between .40 and .50 and for their weighted sum is about .60. The regression of intelligence on criterion or vice versa is distinctly non-linear, for Blakeman's formula gives values of 4.9 and 5.9 respectively. It would thus be hazardous to attempt to include intelligence in a regression equation of higher order. These facts indicate the advantage in as specialized a vocation as agricultural engineering of devising tests for the specific purpose in question.

INTERPRETATION.

If the tests above described are given to an individual for purposes of vocational guidance or selection it is necessary to interpret the scores obtained. This may be done in various ways. The sum of the test scores weighted by the coefficients of the independent variables in the regression equation must of course be computed. This sum may then be taken as the most probable vocational ability of the individual and interpreted, relative to the norms, by considering the standard error of estimate. For instance a normal frequency curve may be drawn with this weighted sum of test scores as the mean and the standard error of estimate as the standard deviation of the curve, or by the use of a table of the probability integral the chances may be expressed of the individual's ability falling within certain assigned limits.

It is perhaps better to develop the interpretation in general terms. It is possible with any correlation coefficient to derive from the table of the probability integral a ten by ten fold table which expresses for any assigned decile in one variable (test) the probable distribution throughout the various deciles of the other variable (vocation). Such a table may be used to predict what per cent of the persons in a given test decile will be in any vocational decile, or to predict for a given individual who scores in a certain test decile his chances of being in any vocational decile. The table below represents such a distribution for a correlation of .60. Some slight adjustments were made to avoid unnecessary decimals. In the columns at the left are given the limits of the deciles of test scores. The first column gives such limits for the weighted sum of the seven tests and the second column for the weighted sum of the two tests in the shorter regression equation.

Test deciles		Vocational deciles									
Seven tests	Two tests	1	2	3	4	5	6	7	8	9	10
.80 to ?	.64 to ?	42	20	13	10	6	4	3	1	1	0
.53 to .79	.42 to .63	20	20	15	13	10	8	6	5	2	1
.33 to .52	.26 to .41	13	15	16	13	12	10	8	7	5	1
.16 to .32	.13 to .25	10	13	13	13	13	11	10	8	6	3
0 to .15	0 to .12	6	10	12	13	13	13	11	10	8	4
0 to –.15	0 to –.12	4	8	10	11	13	13	13	12	10	6
–.16 to –.32	–.13 to –.25	3	6	8	10	11	13	13	13	13	10
–.33 to –.52	–.26 to –.41	1	5	7	8	10	12	13	16	15	13
–.53 to –.79	–.42 to –.63	1	2	5	6	8	10	13	15	20	20
–.80 to –?	–.64 to –?	0	1	1	3	4	6	10	13	20	42

For example suppose an individual scores .65 in the sum of the seven tests. This places him in the second highest decile. Then the chances are 20 out of 100 that he will be in the highest decile in the vocation, 20 out of 100 that he will be in the next highest vocational decile, 15 that he will be in the third decile, 13 that he will be in the fourth, 10 in the fifth, etc. By a decile of vocational ability is meant of course ability like that possessed by the given tenth of the group of students on whom the tests were standardized. A general table like the above is preferable to the scatter plot of the final R of multiple correlation between weighted test scores and criterion because it gives smaller increments.

The above example is typical of the use of the tests for vocational guidance. It makes it possible to tell the student something about his chances of success in this type of work If the tests are to be used, on the other hand, for vocational selection it is necessary to select on the basis of the above distribution table a critical test score and admit to the agricultural engineering course only those who exceed that score. The probable calibre of students admitted or rejected on the basis of such critical score can be determined from the table. If, for example, those with a score algebraically less than–.80 are rejected the expectation is that none of the highest tenth of vocational ability has been rejected but that 42 per cent of those rejected are in the lowest tenth of vocational ability.

Summary

Mental tests were devised for the purpose of predicting ability in agricultural engineering and were standardized upon 43 students specializing in that subject. The tests were given in two installments by means of mimeographed blanks, using the group method. Eighteen test variables were evaluated by means of three instructors' ratings. These ratings were made on a linear scale, reduced to the ratio of x to *sigma* for each instructor and these converted measures averaged for each subject. The final selection of tests was based on inspection of scatter plots for the 18 test variables against the vocational variable. Eight tests were retained, correlated with the criterion, intercorrelated and the regression equation derived.

The correlation between the criterion and the sum of the eight tests weighted according to this regression equation (one weighting was zero) was about .60. However two tests carried most of the load and when weighted according to a regression equation in three variables their sum correlated with the criterion almost as highly as did the eight. Intelligence, on the other hand, had a correlation of only .16 with the vocational variable, which gave further justification of the effort involved in devising special ability tests.

For interpreting the results of the tests when given for purposes of vocational guidance or selection a ten by ten fold table for the correlation of .60 and its standard error of estimate was derived from the probability integral. Knowing an individual's weighted test score it is possible with such a table to state the probability of his attaining any assigned degree of ability in agricultural engineering.

NOTES AND NEWS

THIRD ANNUAL EDUCATIONAL CONFERENCE CONDUCTED BY THE COLLEGE OF EDUCATION, OHIO STATE UNIVERSITY, ON APRIL 5, 6 AND 7.

Nearly 3,000 Ohio educators took advantage of the opportunity to attend the Third Annaual Educational Conference which the College of Education, Ohio State University, conducted at Columbus on April 5, 6 and 7. This attendance marked a decided increase over that of 1922 and emphasized, quite graphically, the increasing interest which the school people of Ohio are showing towards these yearly conferences.

Three general meetings, 24 sectional meetings, and several exhibits were lumped into the three-day gathering as the formal aspects of the conference. In addition to this array, however, there were many informal luncheons and dinners of various groups which were so successfully attended and so thoroughly appreciated that they threaten to become a regular feature of the conference.

The keynote of the conference was that of "Objectives in Education" and all of the meetings were guided by this in their discussions. The summary of the meetings which will be published at an early date are expected, therefore, to be a valuable contribution to the literature on educational objectives since the discussions touched on all phases of education as well as on the field as a whole.

One of the interesting sectional meetings was that conducted by Dr. H. H. Goddard on Clinical Psychology. Dr. Goddard presided at the meetings and conducted a Clinic on the examination and diagnosis of children. Children from the Columbus schools were used as subjects. Mabel R. Fernald, director of the vocation bureau, Cincinnati, spoke at this meeting on "The Service a Clinical Psychologist Can Render the Public School Teacher," and the "Value of the School Psychologist in Personal Work With Pupils" was discussed by H. Austin Aikins, Western Reserve University, Cleveland.

Other sectional meetings which were held were: biological science, commercial education, educational and intelligence tests, educational and intelligence tests demonstration, European history, high school principals' association, Latin, mathematics, parent-teacher's association, teacher training, American history, art, city superintendents, county superintendents, elementary principals, elementary teachers, English, home economics, industrial and vocational education, junior high school principals, modern language, non-biological science and physical education.

One of the attractive features of the Conference was the array of speakers of national reputation who spoke at the different meetings. Among these were: W. C. Bagley and S. S. Colvin, Teachers College, Columbia University; C. H. Judd and Algernon Coleman, University of Chicago; L. V. Koos, University of Minnesota; G. A. Works, *Cornell University; J. C. Miller, Indiana University;* and E. W. Barnhart and F. W. Ballou of Washington, D. C.

One of those chiefly responsible for the Conference has expressed himself as follows: "The remarkable attendance is most encouraging to us in this undertaking to make the Annual Educational Conference a vital factor in the life of the Ohio school system. And the concentration of all the meetings on the problem of selecting clear-cut objectives has meant a clarifying of the aims which the Ohio school people are using as a working basis in the moulding of this system.

"We were especially gratified this year by the participation in our educational program by people whose interest is not primarily educational, such as the business men who aided in the discussion before the commercial education and the industrial and vocational education groups, and too, the meeting conducted for the parent-teachers' association. This gives us a definite point of contact with the people who are supporting the school system and their co-operation makes it easier for us to plan and conduct our work with greater effectiveness."

BOOK REVIEWS

KNIGHT DUNLAP, *The Elements of Scientific Psychology.* St. Louis, C. V. Mosby Co. 1922. 368 pp. Price, $3.50.

In this book Prof. Dunlap has attempted to sift from the great mass of psychological material that essential residue which will be needed by the student of scientific psychology when he comes to deal with the world of "real objects and real activities." While in no sense an attempt at applied psychology it is a very decided attempt to present the facts of psychology in such a way that their application will be possible. The book is intended primarily for college students, but the author assumes that it will also be found useful to the general reader who desires to know the fundamentals of psychology but has no teacher.

The general content of the book may be inferred from the following resume: An introduction dealing with the nature and divisions of psychology; four chapters dealing with sense perception, the senses, and sensory qualities; two chapters on sensory relations and sensory measurements; two chapters on thinking and thought processes; one chapter on the bodily mechanism; two chapters on reactions and reaction types; two chapters on perception; one chapter on affective processes; and one chapter on the empirical self. Two appendices complete the book; the first deals with mental deficiency and mental disease, the second with choice references grouped under certain general divisions.

The following views gleaned from various parts of the text illustrate the author's typical attitudes: Science must be scrupulously exact in the use of its terms, and must furnish experimental proof for all accepted hypotheses. Anecdotal evidence is the bane of pseudo science. The term *sensation* is replaced by the term *senticendum* as being less confusing. Stimuli are classified as physical, unresolved (without mathematical expression,) and inadequate (unusual.) All sapid substances must be soluble "in water." Classifications of odors based upon the method of exhaustion are futile, a more hopeful method will follow more nearly the chemical method of Ramsay. Color vision is based on a slight modification of the Young-Helmholtz theory, the elementary colors of which are red, blue, and chlor (yellow-green.) The pigmental cells of the retina protect the receptors from too great stimulation in strong illumination. Pitch is the extensity character of tones, and depends upon the degree to which the entire tectorial membrane is excited. Pitch differences at first impreceptible are readily preceptible after sufficient practice. The ampullar and vestibular receptors of the ear have no true sensory function, but give rise only to reflexes in skeletal muscles. Weber's law is responsible for the failure to determine many much needed constants of range and threshold values. The term *conditioned reflexed* as frequently used is a misleading term. The contrast between instinct and habit has been overstressed in the past. Intelligence is a term having many meanings, and as far as it relates to intelligence tests may be defined as "whatever intelligence tests measure." Sound localizations in the median plane of the body are made possible by timbre alone. The semi-circular canals are of doubtful function in equilibration. The eye is blind during the saccadic movements of the eye, as in reading. Meanings do not essentially depend upon muscular movements. Association is only a special case of habit formation, and occurs in exactly the same way as does all serial connections.

Throughout the book the author is vigorous and stimulating; his language is not involved and one cannot avoid catching his viewpoint whether he agrees with it or not. While the author has given credit to all authorities cited he has not done so in a way which would permit the reader to refer to the original sources without considerable difficulty. This omission might be a serious handicap to the general reader who knows little of the psychological literature, in fact the serious student of psychology would feel a little more at ease were the opposing views set aside with more opportunity to investigate the sources of authority. All in all the book is well worth reading because it is so different from the usual presentation, if for no other reason.

Ohio University C. H. GROWDON.

ROBERT MACDOUGALL, PH. D., *The General Problems of Psychology*, The New York University Press, New York City, 1922, pp. 464.

That applied science must be grounded in solid general scientific principles is obvious. Almost as obvious is it too, that general science must be based upon satisfactory methodological principles, although this latter situation is frequently overlooked. Too much by far of our scientific work is conceived and executed with faint or no regard to the indispensable methodological background. In consequence our science is not nearly so free from serious defects as it might be.

In psychology especially must the methodological background be constantly kept in mind, for not only is psychology concerned with phenomena which are the most difficult of any to investigate, but also psychological science is, more than any other, a direct derivative from more elementary ways of thinking.

To the thoroughgoing analysis of the methodological perspective of psychology Professor MacDougall devotes the present volume. Very clearly does he point out how serious and obstructive are the misconceptions which arise when psychologists are not fully oriented with respect to the nature and purpose of the conceptions guiding their work. "The individual may be an acute observer or a patient scholar, but, if the guiding assumptions of his work be mistaken, the force of every gift he possesses further subverts the purpose of research and hinders progress. The work done under such conditions is not merely negative. It is an obstacle set in the way of knowledge, whose seriousness is proportionate to its bulk."

The more positive results derived from a thorough acquaintance with the methodological presuppositions which the psychologist uses, not only make for rapidity and security of scientific advance, but also insure the avoidance of insidious misconceptions. Professor MacDougall in the following quotation enumerates some of these misconceptions which have been made in the past. "The mind's activity is not determined by a sufficient body of inductions to make its speculations rational. Fantastic hypotheses are shaped. Gross perversions of the meaning and scope of science arise. Methaphysical and theological concepts invade its territory. The thinker turns from the patient exploration of fact to spin deductions from a system of a priori definitions. Fact and myth are interwoven in a new unitary complex. The natural connections of cause and effect are supplemented by essences and forms, magic and miracle, as method shifts unpredictably between theological interpretation and explanation in terms of natural law."

The treatment which Professor MacDougall gives to the general problems of psychology is at once comprehensive and thorough. He discusses not only the relation of psychology to knowledge in general and to the arts of life, but the field, the limits, the methods, data, and specific problems of psychology. It is not necessary to be in full agreement with Professor MacDougall to appreciate that this book is of considerable importance and well worth the perusal of every serious student of psychology.

A. A. ROBACK, *Behaviorism and Psychology*, University Bookstore, Cambridge, 1923, pp. 284.

As its title indicates "this book has had for its purpose the confutation of the claim of behaviorism to the title of psychology." Although the author summarizes his destructive efforts in the following statement, that statement does not convey in full force the severely critical judgment he really entertains of the behaviorists.

"I have tried to show that (1) the concept of behavior is too general to afford us the foundation for a separate science, (2) behaviorism, in its restricted sense, seeks arbitrarily to dominate a field which is not, *ex hypthesi*, under its jurisdiction by bowing out of court everything which does not fit into its narrow frame work, and (4) behaviorism does not readily lend itself to application."

How then is the work of destroying such a "general" and worthless movement as behaviorism carried out? Briefly, by putting all the behaviorists into one group and attacking all through what the author considers the weakest point in each of the very few that he slightly examines.

Similarly, behaviorism is deprived of all applicational value by the simple method of assuming that only mentalism is psychology. In consequence of this fact, it is asserted that if one wishes to handle situations involving intention, belief, intelligence, ideas and emotions, or other such psychological facts, behaviorism cannot be of any service at all.

Indiana University. J. R. KANTOR.

EDWARD ALLSWORTH ROSS, *The Social Trend*. Century Co., New York, 1922, pp. xxii and 235.

War, prohibition, "philanthropy with strings," women in the home and in a man-made world, migration, the vanished frontier, the birth and death rates, a legal dismissal wage: such are some of the topics considered in this collection of papers. It were vain to call it a book.

Written with all the very great attractiveness of Professor Ross' style, fearless, filled with interesting views, the volume deserves a wide reading.

Perhaps none of these papers is more acute than the one on "The Legal Profession from the Social Point of View" or the following one on "The Conscience of the Expert." It is shown that the skill of able lawyers retained by the wealthier, and when wrong more reckless side, tends to whittle down the defenses of the law. The lawyer is in large measure engaged in an anti-social undertaking. Compare with him the true expert, the product of modern training in scientific method, possessing the scientists love for truth and you have almost a complete antithesis. The one aims to establish a point of view, his side, whether true or false, against the assaults of his brother lawyers; the other to wrest the truth from nature and report it. Society should and will make more use of the scientific expert, not so much because he has a special method or special knowledge, but because he has a conscience sensitive to truth. So be it.

All of these papers deserve inclusion in the collection and their very variety serves to indicate the diversified social trends of a complex social order.

B. F. HAUGHT, *The Interrelation of Some Higher Learning Processes.* Psychological Monographs, Whole No. 139. Princeton, N. J. Psychological Review Co. 1921. pp. vi and 69

This is a study, with very complete statistical discussion, of some so-called higher learning processes, "rational learning," "checker puzzle," "Tait labyrinth" and others, and a comparison of the results of these with the results from the Stanford-Binet. The intercorrelations are low and conflicting and the conclusions are therefore adverse to the Spearman "Two Factor Theory" of intelligence. Are logical conditions and factors in learning on the whole less determinative than such matters as emotional stability or disturbance, perseverance, self-control and so forth?

The Professional Education of Teachers in Cleveland. Cleveland, Ohio. Division of Publications, The Board of Education. 1922. pp. x and 92. Price $ .50

We have here the report of the Educational Commission which was appointed to study the subject named in the title. This Commission was composed of Professors W. C. Bagley, J. W. Withers and Geo. C. Chambers. The pamphlet discusses and makes recommendations on the affiliation between the Cleveland School of Education and the Western Reserve University. It considers admission, organization, curricula, finance and other problems and gives in appendices a number of documents which show requirements and organization of this important joint venture in the pre-service and in-service training of teachers for a great city. The report is an important contribution to current discussion of teacher training.

R. G. JONES, *Survey of Department of Instruction of the Cleveland Public Schools.* Cleveland, Ohio. Division of Publications, Board of Education. 1922. pp. 29. Price $ .50

This is the first of a series of continuous surveys to be conducted by the Department of Research of the Cleveland Schools. It provides a brief analysis of the educational operation of the school system. There are diagrams and tables on enrollment, promotion, age-grade conditions, achievements in individual subjects, the work of the psychological clinics, etc. The material was prepared under the direction of Dr. W. W. Theisen.

ARTHUR I. GATES, *The Psychology of Reading and Spelling, with special reference to disability.* Teachers College, Columbia University Contributions to Education, No. 129. New York, Teachers College, Columbia University, 1922, pp. viii & 108.

It is sixteen years since Huey published his excellent summary of the psychology and pedagogy of reading. Since that time a great quantity of experimental and clinical work has been done in the field of reading, speech and spelling; and in consequence, it is becoming apparent that the difficulties involved are only now slowly emerging. The solutions are still to be sought.

The present investigation was carried on at the Scarbrough School, Scarborough, New York. Dr. Gates had the good fortune to work with a group of pupils of normal or superior intelligence but who yet manifested the reading or spelling difficulties which form the subject of study. A number of group tests of general mental ability, of achievement in school subjects, of responses to a variety of visual and auditory stimuli and of several sensory mechanisms and motor responses were given to a group of one hundred pupils in grades III to VIII. Detailed individual tests of mental ability, as the Stanford-Binet, and of specialized abilities in pronounciation, word perception, etc., were also given. The investigation was carried out with considerable detail and great care.

In a short review it is impossible to make any adequate report of the study. In the first part of the volume it is shown that there is an ability common to reading, the recognition of correct spelling, proof reading, and the ability to detect differences between pairs of similar words. It would seem that both for prediction and for diagnosis the value of any one of these tests is *minimized by the fact that they are applicable only through the* use of material which has already been learned; and that the *worth of showing that there are psychological elements common* to reading and proof reading, for example, is not great.

It appears that the most fruitful portions of the study are the short reports of individual difficulties in reading and spelling. (See pp. 41-42 and 66-67, etc.) It may at least be seriously questioned whether we will ever get very much further with group studies of reading defectives until after we have made more detailed and critical studies of a large number of individual cases and defects.

An interesting conclusion, in view of the work of Huey and his successors, is the statement that inappropriate eye movements are not the cause but only a symptom of poor reading, and even then, not the symptom most easily observed. (p. 57)

Concluding chapters deal with an analysis of abilities involved in spelling and present criticism of various theories of the causes of backwardness in reading and spelling.

Perhaps the most important conclusion of the study—and it is certainly an important one—is that presented in the following sentences: "A case of inability to read affords, frequently, a tangle of difficulties that experts from several professional fields working together may be unable to disentangle. Such a situation *portrays clearly the need of a new group of specialists who will* make the solution of such problems their main work." (p. 106)

Ohio University H. G. Good.

HENRY E. JACKSON, *Robinson Crusoe, Social Engineer*. E. P. Dutton, New York. 1922. pp. 301. Bibliography of more than 60 titles.

The author divides discussion into three parts. In part I he describes at length the character of Robinson Crusoe. He sees in him a typical Anglo-Saxon actuated by the gospel of work and of courage. Crusoe is represented as meeting numerous discouragements from which he turns unshaken and unafraid.

In part II the author presents certain essential characteristics of the industrial situation as it is found today. The author maintains that the source of the difficulty is found in the fact that the principles of feudalism control industry today. Industry fails to recognize the equality of man a principle which has been recognized for more than a hundred years in politics. He declares that industry in this particular is a century and a half behind politics. Politically men are equal but industrially the old relation of lord and vassal maintains.

In part III the author proposes a solution for the industrial problem. In brief he declares that the solution will be reached when a new attitude of mind has taken possession of those who supply capital and those who supply labor. This new attitude of mind must recognize the principle of manhood—the right of each individual to a proper return upon his investment whether this be in terms of capital or labor. In a word it is a plea for industrial democracy.

The book is full of optimism and faith in the future. The author does not attempt to list successful applications of the new doctrine. He is concerned only with the presentation of a principle of conduct. His discussion is clear and compels consideration on the part of the reader.

FRANK G. PICKELL, *Junior High School Mathematics*. Cleveland, Board of Education. 1922.

This is a course of study and syllabus for mathematics in the Junior High School in Cleveland. In this syllabus material suitable for mathematics in junior high schools has been organized along lines independent of the traditional organization. In the 7th grade graphs, fundamental processes, business forms and accounts, equations, and percentage with its applications, are outlined. In the 8th grade the fundamental processes of arithmetic are further considered and advanced work in the application of percentage as well as intuitive geometry are treated. In the 9th grade emphasis is placed upon algebra. In the treatment of algebra in this grade one important change from the traditional sequence has been made; factoring and simultaneous equations change places. "This change has been made first to stress at the outset the equation as the most valuable tool in the solution of problems and second to prevent the waste of time to the 9B pupils."

KLINE-CAREY. *The Revised Measuring Scale of Free-hand Drawing*. The John Hopkins Press. 1923. Price $ .60. Baltimore, Md.

This is a revision of the authors' earlier scale for measuring free-hand drawing. The revision presents a scale for measuring achievement in the drawing from memory of distinct types of objects. These objects are: house, rabbit, figure in action, and tree. This has made the merit of furnishing an instrument for measuring achievement along special lines.

Ohio University. W. L. GARD.

WILLIAM T. ROOT, JR. *A Socio-Psychological Study of Fifty-three Supernormal Children.* Psychological Monographs Vol. XXXIX, Psychological Review Company, Princeton, N. J., and Lancaster, Pa.

This study of supernormal children was undertaken because the author felt the importance of the Superior Child to the state and because the lack of information concerning the factors making for superiority give the subject peculiar interest. The aims of the study are mainly to compare the behavior of the superior child with that of the normal child; to note the individual differences of the superior mind in methods of procedure, imagery, associative and mnemonic devices; the qualitive character of their responses; and to try to discover the significance of definite systems of training, by incidental education, and the social milieu.

The children for this research were selected from the city of Los Angeles. The preliminary selection was made thru the schools, newspaper accounts, etc. The final selection was based upon the results of Binet Tests scores and special ability along some line. Several tests were used in the experiment, such as Binet Letter-Square, Trabue Completion Test, Healy-Fernald Construction Puzzle, Proverb Test, etc., and the different phases of mentality studied. Not only the Psychological test given to discover the mental ability of each child were studied, but the social histories were thoroughly investigated. The effects of home life and training upon the mental development of each child were carefully considered.

The author's conclusions are most interesting and valuable to the research student in this field. Mr. Root has reserved no time and energy in stating clearly what he discovered in this work with bright minds.

Ohio University LAVINIA WARNER.

THOMAS VERNER MOORE. *PERCY BYSSHE SHELLEY, An Introduction to the Study of Character.* The Psychological Monograph, Vol. XXXI, No. 2. Whole No. 14, 1922.

This monograph is an important and stimulating endeavor to apply the literary-historical method of psychoanalysis to Shelley.

In the author's analysis of the mentality of the poet he finds, first of all, the "precox elements of negativism and defense." He is firmly of the opinion that Shelley's plan of life, or its chief driving force consists of an "affective drive for an object of sensuous love" and an intellectual drive colored by protest.

In the development of his thesis the writer seems to insist upon the idea that the prime dominating motive force of Shelly's life was his yearning for a beautiful woman who could unreservedly appreciate him and his poetical capabilities. In support of this view he draws chiefly upon *Alastor*. However, a question of once arises as to whether this poem should receive the interpretation that Mr. Moore appears to give it. Critics have largely agreed that in this significantly autobiographical poem the Poet's quest of the lovely maiden may be any poet's quest of ideal beauty—such beauty as is celebrated in the *Hymn to Intellectual Beauty*. And although one readily concedes that Emilia Viviani, more or less directly inspired *Epipsychidion*, one much prefers to give the later poem a spiritual rather than a literal signification. Shelley here emphasizes a fact so often expressed by him that a poet must ever seek the elusive beauty of life and poetry and that when he attains to such only disappointment and failure must ensue. The ideality and unrealized end of the pursuit are the agencies impelling a poet. Had the Poet in *Alastor* attained the alluring object of his love he would have experienced defeat in the very fact of his realization; only the continuous striving toward her brings richness or reward.

As a potent influence in Shelley's life Mr. Moore cites the poet's relationship with Harriet Westbrook. He finds little justification for Shelley's conduct. It is to be admitted that even the most enthusiastic admirers of Shelly cannot make an altogether satisfactory brief for him. However, it is gratifying to the lovers of Shelley to note William Sharp's assurance that Shelly's "sorrow and life-long regret for the malfortunate outcome of his action were genuine."

Mr. Moore well points out the element of protest in the poet's character. But one questions whether this factor is wholly, or even largely, due to the "father complex." One is inclined to hold that the Sion House experience, more than any other, enforced, if not implanted, Shelley's hatred of tyranny. Unquestionably there is some difference of opinion as to the poet's home relationships; but all agree as to the extreme distress he suffered at Sion House. Mr. Moore draws upon the Prometheus myth to illustrate his convictions in this instance.

The Prometheus theme, a very popular one in literature, perhaps cannot be summoned to demonstrate anything peculiar to Shelley. True the poet employs it as a medium for his expression of the overthrow of all forms of tyranny and his installation of the Golden Age. Demogorgon, a name which in the drama no longer carries any connotation of magic, is by the poet defined as "Eternity." Shelley has not identified him with mere Necessity because he too strongly personalizes the character; for instance, the closing pronouncement of the last act. The poet's dream reaches its most vivid delineation in the eternal union of Love and Humanity.

Chapter IX of Mr. Moore's article contains a brief analytical summary of the Shelley profile. It is reasonably comprehensive. As has just been indicated one is inclined to differ from the latter test against all authority rooted unconsciously in the hatred of portion of the second predication, namely, "The bold, violent pro-

Books and pamphlets for review should be sent to James P. his father." Perhaps one might suggest that greater emphasis should have been placed on Shelley's generosity of soul. The philosophical-religious aspect of the poet's mind has not been noted. On an earlier page Mr. Moore refers to Shelley's homosexuality. Fortunately this is omitted in the profile. It would be difficult to detect this element in Shelley's mental constitution.

Mr. Moore's careful and scholarly work is, in most respects, to be warmly commended. He has given us an effective attempt to ascertain the mental traits that most palpably have revealed themselves in the operations of Shelley's mind.

Ohio University HIRAM R. WILSON.

The Honesty Book. Prepared and published by The National Honesty Bureau, 115 Broadway, New York.

A textbook for schools written by business men instead of by pedagogues, a good idea. Hitherto the most recent illustration in schoolbooks of business honor have been taken from such remote examples as Benjamin Franklin and Peter Cooper. This little book begins with a talk about a modern office-boy, tells how a young man today gets a bond and it is replete with references to the business practices of our own time. The National Surety Company has originated the plan of reviving the commandment. "Thou shalt not steal," in the schools and has had the stories and methods tried out successfully in real schools by teachers with actual pupils. The stories are crisp, pointed, unconventional and avoid the goody-goody.

The volume is an interesting attempt to demonstrate how a particular virtue may be inculcated. Its suggestions give promise of being helpful to teachers in teaching other virtues. The book and the campaign of which it is the instrument are to be heartily commended.

Ohio University JAMES P. PORTER.

NEW BOOKS AND PAMPHLETS RECEIVED*

Porter, Department of Psychology, Ohio University, Athens, Ohio.

Abel, J. F. *State Aid to Weak Schools.* Rural School Leaflet No. 7, Department of the Interior, Washington, D. C. 12 pp.

Abbott, Julia Wade. *Kindergartens Past and Present.* Kindergarten Circular No. 11, Department of the Interior, Bureau of Education, Washington, D. C. 5 pp.

Applied Science Course for High School Boys. Home Economics Circular No. 16, November, 1922, Department of the Interior, Washington, D. C. 7 pp.

Athearn, Walter S. *The Indiana Survey of Religious Education.* Committee on Social and Religious Surveys, 111 Fifth Ave., New York. Price 25 cents. 40 pp.

Bach, Teresa. *Education in Czechoslovakia.* Bulletin No. 39, 1922, Bureau of Education, Department of the Interior, Washington, D. C. 26 pp.

Bach, Teresa. *Education in Poland.* Bulletin No. 41, 1922, Bureau of Education, Department of the Interior, Washington, D. C. 21 pp.

Barnes, Earl. *How the Kindergarten Makes Americans.* Kindergarten Circular No. 9, Bureau of Education, Department of the Interior, Washington, D. C. 6pp.

Barrows, Alice. *First National Conference on the Work-Study-Play or Platoon Plan.* Bulletin No. 35, 1922, Bureau of Education, Department of the Interior. Washington, D. C. 16pp.

Blaisdell, J. Harper. *The Menace of Syphilis to the Clean Living Public.* Massachusetts Society for Mental Hygiene. 15 Ashburton Place, Boston, Mass. 18 pp.

Blauch, L. E. *Statistics of Land-Grant Colleges, Year Ended June 30, 1921.* Bulletin No. 34, 1922, Bureau of Education, Department of the Interior, Washington, D. C. 67 pp.

Bousfield, Paul. *The Omnipotent Self.* A Study in Self-Deception and Self-Cure. E. P. Dutton & Company, 681 Fifth Avenue, New York. Price $2.00. 193 pp.

Budget of The State Board of Charities and Corrections. Jefferson City, Mo. Vol. 24, No. 9, December, 1922. 8 pp.

Dearborn, Walter F., *Edward A. Lincoln* and *Edwin A Shaw. Standard Education Tests in the Elementary Training Schools of Missouri.* Harvard Monographs in Education, Whole No. 3, Series 1, No. 3. The Graduate School of Education, Harvard University, Cambridge, Mass. 90 pp.

DuBois, Florence *and* H. R. Bonner. *Statistics of State Schools Systems 1919-20.* Bulletin No. 29, 1922, Department of the Interior, Washington, D. C. 68 pp.

Educacion, Revista Mensual, Vol. 1, Nums. 5 y 6. Apartado Postal, 1954, Mexico, D. F. 380 pp.

Educational Directory 1922-23. Bulletin No. 50, 1922, Bureau of Education, Department of the Interior, Washington, D. C. 179 pp.

Eugenics, Genetics and the Family Vol I. and *Eugenics in Race and State,* Vol II. Scientific Papers of the Second International Congress of Eugenics. Williams & Wilkins Company, Baltimore, Md. Vol. I, 438 pp. Vol. II, 472 pp.

---

* Mention here does not preclude further comment.

EVANS, HENRY R. *Educational Boards and Foundations, 1920-1922.* Bulletin No. 38, 1922, Bureau of Education, Department of the Interior, Washington, D. C. 11 pp.

FOX, FLORENCE C. *Major Projects in Elementary Schools.* Bulletin No. 36, 1921, Bureau of Education, Department of the Interior, Washington, D. C. 43 pp.

GAULT, ROBERT H. *Social Psychology.* Henry Holt & Company, New York City, 336 pp.

HAUGHT, B. F. *The Interrelation of Some Higher Learning Processes.* Psychological Monographs, Vol. XXX., No. 6, Psychological Review Company, Princeton, N. J. 71 pp.

High School Buildings and Grounds. Bulletin No. 23, 1922, Bureau of Education, Department of the Interior, Washington, D. C. 49 pp.

HOKE, ROY EDWARD. *Improvement of Speed and Accuracy in Typewriting.* Johns Hopkins University Studies in Education, No. 7. Johns Hopkins Press, Baltimore, Md. Price 75 cents. 41 pp.

HOOD, WILLIAM R. *Some Important School Legislation 1921 and 1922.* Bulletin No. 43, 1922, Department of the Interior, Washington, D. C. 27 pp.

KELLY, FRED C. *Business Profits and Human Nature.* G. P. Putnam's Sons, New York City. 279 pp.

Kline-Carey Measuring Scale for Free-Hand Drawing. Johns Hopkins University Studies in Education, No. 5a. John Hopkins Press, Baltimore, Md. Price 60 cents.

LATHROP, EDITH A. *Essential Features of Laws Concerning Transportation of Pupils.* Rural School Leaflet No. 8, Department of the Interior, Washington, D. C. 7 pp.

LATHROP, EDITH A. *Organization of the One-Teacher School.* Rural School Leaflet No. 10, February 1923, Department of the Interior, Washington, D. C. 12 pp.

Length of School Sessions in Grades One and Two. City School Leaflet No. 6, Department of the Interior, Washington, D. C. 4 pp.

List of References on the Project Method in Education. Library Leaflet No. 17, Department of the Interior, Washington, D. C. 9 pp.

LOOSMORE, W. CHARLES. *The Gain of Personality.* E. P. Dutton & Company, 681 Fifth Avenue, New York. Price $2.50. 238 pp.

McKERROW, JAMES CLARK. *The Appearance of Mind.* Longmans, Green & Company, 55 Fifth Avenue, New York. Price $2.00. 120 pp.

Marine Biological Laboratory Annual Announcement. Woods Hole, Mass. 36 pp.

Measurement of Achievement in Shorthand. Johns Hopkins University Studies in Education, No. 6, Price $1.50. Johns Hopkins Press, Baltimore, Md. 118 pp.

MILLER, WILLIAM L. *County Almshouses and Jails in Missouri.*

Bi-Monthly Bulletin, Vol. 24, No. 7, State Board of Charities and Corrections, Jefferson City, Mo. 7 pp.

Missouri Reformitory Report and Recommendations. State Board of Charities and Corrections Bi-Monthly Bulletin, Vol. 24, No. 8. Jefferson City, Mo. 12 pp.

Mitchell, T. W. *Medical Psychology and Psychical Research.* E. P. Dutton & Company, New York. Price $3.00. 244 pp.

Nuestros Problemas Medico Sociales. Ministerio de Fomento, Direccion de Salubridad. Lima, Peru. 18 pp.

O'Shea, M. Vincent. *Tobacco and Mental Efficiency.* The Macmillan Company, New York. 258 pp.

Phil. Al. Phil. .*The Soul of the State* or *The Know Thyself.* Vol. I. Athanasios Papaspyrou, Stoa Simopoulou, Odos Leka, Athens, Greece. 140 pp.

Publications Available November, 1922. Department of the Interior, Bureau of Education, Washington, D. C. 20 pp.

Research Bureaus in City School Systems. City School Leaflet No. 5, Department of the Interior, Washington, D. C. 23 pp.

Report of the Massachusetts Society for Mental Hygiene. Publication No. 7. 1132 Kimball Building, 18 Tremount Street, Boston, Mass. 16 pp.

Ross, Edward A. *The Social Trend.* The Century Company, New York. 235 pp.

Report of Secretary of War on Conference on Training for Citizenship and National Defense. Superintendent of Documents, Government Printing Office, Washington, D. C. Price 5 cents per copy. 36 pp.

School Life. Vol. VIII, No. 4, Department of the Interior, Washington, D. C.

Sears, Jesse Brundage. *Philanthropy in the History of American Higher Education.* Bulletin No. 26, 1922, Department of the Interior, Washington, D. C. 112 pp.

Shaw, Edwin A. and Edward A. Lincoln. *A Comparison of The Intelligence and Training of School Children in a Massachusetts Town.* The Graduate School of Education, Harvard University, Cambridge, Mass. 49 pp.

Statistics of Public High Schools, 1919-1920. Bulletin No. 37, 1922, Department of the Interior, Washington, D. C. 41 pp.

Strong, Edward K. and Richard S. Uhrbrock. *Job Analysis and The Curriculum.* Williams & Wilkins Company, Baltimore, Md. 146 pp.

Studies in Citizenship for Citizens Military Training Camps. United States Army Training Manual No. 2. War Department, the Adjutant-General's Office, Washington, D. C. 40 pp.

Wells, F. L. *The Status of "Clinical" Psychology.* Reprint and Circular Series of the National Research Council, No. 31. Psychological Laboratory, Boston Psychopathic Hospital, Boston, Mass. Price 20 cents. 12 pp.

Zook, George F. *Accredited Higher Institutions.* Bulleton No. 30, 1922, Department of the Interior, Washington, D. C. 106 pp.

THE
JOURNAL OF APPLIED PSYCHOLOGY
VOL. VII SEPTEMBER, 1923. No. 3

THE RETURN COUPON AS A MEASURE OF ADVERTISING EFFICIENCY

BY A. T. POFFENBERGER

Columbia University

When the psychologist offers a device for measuring the effectiveness of advertisements, such as the Order of Merit Method, or one of the Attention or Memory methods, it should be evaluated not in comparison with some ideal method but rather with the other actual methods in use with which it would have to compete. In investigating current methods of measuring advertising efficiency the writer was surprised to discover the reliance which is at present placed upon the return coupon as a measure of the value not only of a particular piece of advertising copy, but of an advertising medium which carries the copy. This reliance seems not to be limited to cases of strictly mail order selling, but is extended to cases where the product may be purchased directly in any locality, and where the return coupon concerns merely samples or descriptive literature. Publications have been known to lose valuable clients because the number of "replies" through that medium do not come up to expectations. There must be many factors determining whether readers of a particular medium will or will not indicate their interest in advertised articles by filling in and mailing a return coupon. The price of a magazine as well as the nature of its contents, determining the character of its readers; its editorial policy such as that of advocating buying from local dealers; the distribution of readers in city and country; all may give clues to the possible attitude of the reader toward the return coupon. In addition to these, there are other more strictly psychological factors involved. These, like most others which determine the human reaction to advertising, can be discovered by painstaking research. This report deals with conditions that are not peculiar to any one medium, but rather with the effectiveness of the return coupon in general. Two of these will be discussed, namely (1) the dependence of effectiveness upon attention value, and (2) the dependence of effectiveness upon ease of using the coupon.

1. A casual survey of any magazine will provide abundant evidence of the lack of attention-getting power in the return coupon. Its size is usually reduced to a minimum; its position is at the extreme bottom of the advertisement or in a lower corner, with none of the customary mechanical devices employed to carry the eye to it; its type is small and plain. It would seem that a reader might well have his interest aroused in the article advertised and yet fail to be stirred to action by the appeal of the coupon, or worse still might fail to notice the coupon at all.

To test the attention value of return coupons in comparison with the advertisements that carry them, a simple experiment was arranged. Six full page advertisements (black and white only) were taken from the October 14 issue of the Saturday Evening Post. Three of these contained return coupons of the usual sort, while the other three had none. This set of six advertisements was presented to 210 persons each of whom was asked to examine them as much as he liked, but more carefully than he would in looking over a magazine. Immediately after this examination was finished, the advertisements were withdrawn and each person was asked to write the names of the articles advertised. The test was made extremely easy by using only six advertisements and testing for knowledge of them immediately after so that as large a proportion of the people tested as possible should remember all of them. That most of the persons did remember all the advertisements may be seen in column 3 of the accompanying table, where 189 of the total of 210 remembered the advertisement making the lowest score and 202 of the total of 210 remembered the one making the highest score. It will be noticed that the six advertisements were nearly equally well remembered.

As soon as the names of the articles advertised had been written, these sheets were collected, and others, containing a printed list of these articles were given to each person with instructions to check those articles whose advertisements had return coupons. The results of this test are given in the table. The first column gives the name of the article advertised, columns 2, 3 and 4 give the number of persons who recalled each advertisement. The scores for men and women are reported separately in columns 2 and 3 and are then combined in column 4. Columns 5, 6 and 7 give the data for the recall of the coupons, 5 and 6 giving the records separately for men and women and 7 combining these into a total score. *The first three advertisements in the table contained return coupons while the last three contained no coupons.* The lower right hand corner of the table—

figures followed by an (*)—represent the cases where advertisements were said to contain coupons when they really did not. Disregarding the records for men and women separately, we find that 30, 55 and 31 persons reported falsely concerning the presence of coupons. These figures indicate the degree to which guessing occurred, and furnish the means for calculating approximately how much guessing occured in the case of advertisements which really did have coupons. This may be done roughly as follows: It will be observed that the last three advertisements were recalled by 592 persons and of these 116 or about 20 per cent said they contained coupons when they did not. The first three advertisements were recalled by 571 persons and of these 212 or about 40 per cent said they contained coupons. But some of these persons must have been guessing and we may assume that about the same proportion guessed here as in the case of the last three advertisements, namely, 20 per cent. Making this correction, it would appear that only about 20 per cent of the people who recalled the advertisements recalled whether they had return coupons. As our correction for guessing is conservative, this 20 per cent is more likely to be too high than too low.

NUMBER OF PERSONS WHO	RECALLED ADVERTISEMENTS			RECALLED COUPONS		
	ADVERTISEMENTS WITH COUPONS					
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Articles Advertised	*Men*	*Women*	*Both*	*Men*	*Women*	*Both*
Fleischman's Yeast	81	109	190	40	68	108
Booth's Sardines	82	110	192	18	31	49
Underwood Bookkeeping Machine	83	106	189	22	33	55
Total	246	325	571	80	132	212
	ADVERTISEMENTS WITHOUT COUPONS					
Styleplus Clothes	88	111	199	14*	16*	30*
Squibb's Products	86	105	191	25*	30*	55*
Apollo Chocolates	91	111	202	12*	19*	31*
Total	265	327	592	51*	65*	116*

Since the ultimate measure of the value of an advertisement is the number of purchasers it creates, the return coupon as an indicator of effectiveness should register more than 20 per cent of the attention and memory value of the whole advertisement. By giving it some of the attention power devoted to the rest of the advertisement, it might be made to register more nearly the value of the advertisement or of the medium in which it is placed.

Certainly if much reliance is to be placed upon the coupon to indicate these values some effort should be made to bring it to the attention of the reader.

The figures in the table showing the records for men and women separately both in the recall of advertisements and in the recall of coupons show only one difference worth noting. In every case the scores for women are higher than for men, a difference that is usually noted in tests of this sort. But the proportion of coupons recalled to advertisements recalled and the proportion of guesses does not differ materially for the two sexes. Likewise, no particular significance is to be attached to the difference in scores of the different advertisements. They were too few in number to enable conclusions to be drawn about the value of types of coupons etc. The high coupon score for the Fleischman advertisement is probably due to the large element of guessing in that case, but this cannot be determined definitely from our records. The influence of difference in size, position and layout of coupons is a matter which would be well worth investigating.

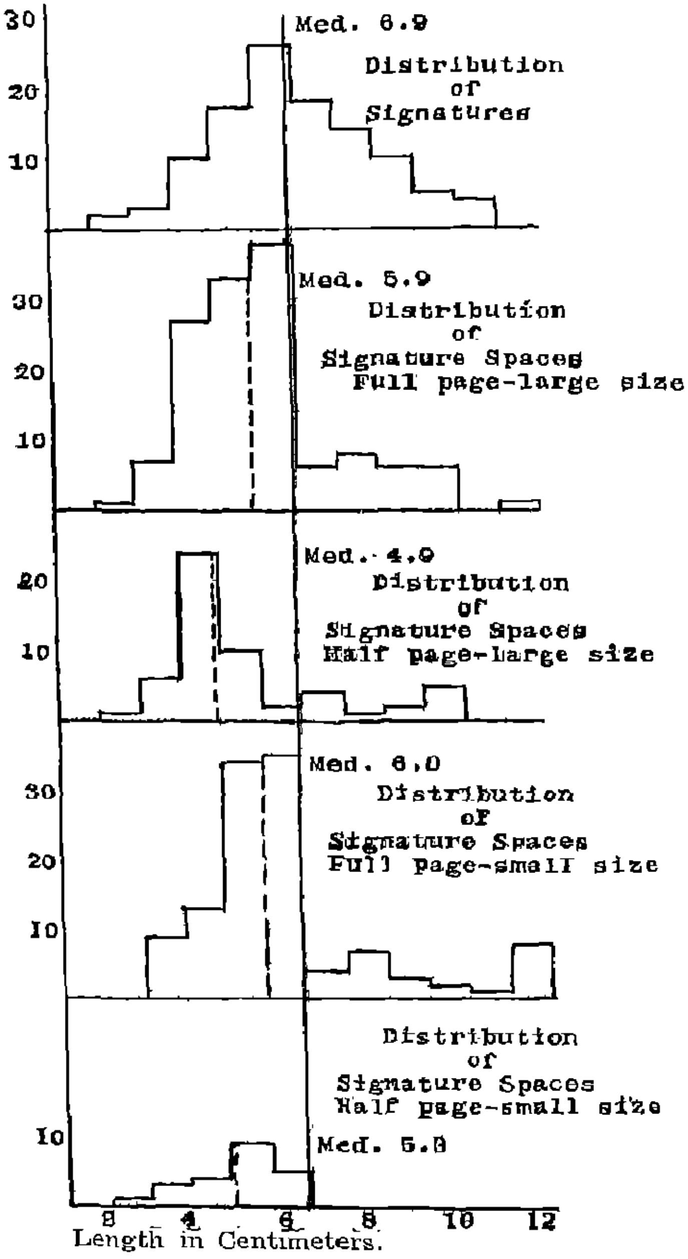
2. The second condition upon which the effectiveness of the return coupon depends is the ease with which it may be used. There are a number of factors that might well be considered here. For example, the paper on which the advertisement is printed may be too smooth to permit of pencil writing and too porous to permit the use of ink without its blotting. Again, if not perforated, the coupon may tear in being removed from the advertisement. These are difficulties probably inherit in the medium and which cannot be remedied. But there is another, which can be more easily handled. This has to do with the amount of space allowed for writing the name and address on the coupon. A glance at a collection of coupons will show that many of them are entirely too short to receive an ordinary signature. Now, it may seem at first a very small matter to expect the reader to reduce his handwriting sufficiently to meet the coupon conditions. Still, it must be remembered that in many cases, if not in the majority of them, the decision to react or not to react by filling in and mailing a coupon hangs on a very slender thread. The resistance which any slight obstacle offers may be just enough to sway the re-action in the negative direction.

A series of measurements was made of a random collection of 110 signatures, and a series of 435 return coupons, some of them from full page and some from half page advertisements in large size periodicals such as *The Ladies Home Journal* and the *Saturday Evening Post;* others were taken from full page and half page advertisements in standard size magazines such as *Scribners* and the *Century Magazine*. The spaces allowed for both the signature and the address were measured but only the former will be discussed. The accompanying chart will show graphically the degree to which the spaces fall short of accommodating the signatures. The horizontal scale for every curve represents the length of the signatures and the signature spaces in centimeters, and the vertical scale represents the number of cases of a given length. The topmost curve shows the distribution of *signature* lengths, ranging from about 2.5 cm., to about 11.5 cm., with the median length at 6.9 cm. The second curve shows the distribution of signature *spaces* in coupons from full page advertisements in large size magazines; the third, in coupons from half page advertisements in large size magazines; the fourth in coupons from full page advertisements in standard magazines; and the fifth in coupons from half page advertisements in standard size magazines. The solid line representing the median length of the signatures is carried down through the other curves, whose own medians are represented by dotted lines. Although there are a few coupons with enough space to accomodate the largest names, the median of the names is considerably larger than the medians of any of the spaces. The relationship may also be expressed in terms of the per cent of the signature *spaces* that equal or exceed in length the median of the signatures. These figures are given below.

	No. Cases	Median Length in Cm.	% Reaching or Exceed. Med. Name
Signatures	109	6.9	20
Full page Ad. Large Size	133	5.9	20
Half page Ad. Large Size	55	4.9	20
Full page Ad. Standard Size	116	6.0	20
Half page Ad. Standard Size	22	5.3	0 (Approx)

If the distributions of the spaces were equal to those of the signatures, there should be 50 per cent of the cases of the former exceeding the median of the latter. Thus in three of the cases, 30 per cent of the names are too large for the spaces, and in the fourth case, they are about all too large to go in the space provided. There appears to be an even greater discrepancy between the *height* of writing space provided and the height of space needed for the signatures, although no actual measurements have been made to support this statement.

That the space allowed for coupon purposes is not limited by the size of the advertisement is clear from the fact that the coupon in the standard magazines tended to be larger than that in the


30
20
10
Med. 6.9
Distribution
of
Signatures
30
20
10
Med. 5.9
Distribution
of
Signature Spaces
Full page-large size
20
10
Med. 4.9
Distribution
of
Signature Spaces
Half page-large size
30
20
10
Med. 6.0
Distribution
of
Signature Spaces
Full page-small size
Distribution
of
Signature Spaces
Half page-small size
10
Med. 5.8
2
4
6
8
10
12
Length in Centimeters.

large size magazine, both for the full and half page advertisements (medians of 6.0 and 5.3 compared with medians of 5.9 and 4.9). The real reason is to be sought in the failure to recognize the importance of providing adequate space. Adults have their writing habits firmly established, so much so that writing a signature or an address is quite automatic. Interference with the smooth flow of the automatic responses when once begun, arouses resistance and an unpleasant feeling-tone. This is illustrated in the well-known Downey tests for character traits in which the subject is asked to reduce, enlarge or otherwise alter the character of his writing, and in which the measure is in terms of resistance encountered in complying with the instructions. Now, it is just this unpleasant feeling-tone which the advertiser tries in so many other ways to avoid, e.g. by the use of beautiful illustrations, beautiful color combinations, graceful border treatments, appropriate type faces, etc. The coupon, if it is to be relied upon or used at all, should certainly not be permitted to defeat its own purpose or the purpose of the whole advertisement by allotting too little space to it.

This brief study points to three conclusions, namely, that the return coupon cannot adequately measure the effectiveness of an advertisement or a medium, except in direct mail selling, because of the complexity of factors on which reaction to it depends; that as used at present it lacks sufficient attention-getting power to attract more than a small portion of the readers of the advertisement; and that in the size in which it most commonly appears it is likely to create an unpleasant feeling-tone because of inadequate space allowed for writing.

FLUCTUATIONS IN MENTAL PRODUCTION WHEN MOTIVATION IS THE MAIN VARIABLE

By F. B. KNIGHT and H. H. REMMERS

GAP BETWEEN POSSIBLE AND ACTUAL PRODUCTION IN COLLEGE STUDENTS

No one knows with useful accuracy just how great is the gap between the potential capacity of a college student and his accomplishment in college courses. There are, no doubt, many students who, if we make a rough use of Franzen's Accomplishment Quotient, possess 100 units of potential capacity and whose accomplishment in college work is appropriately 100 units. These are efficient students, for there is a negligible gap between potential power and mental production. On an invested capital of ability the yield is satisfactory.

There are other students who possess say 100 units of capacity, but whose mental production is 80 units, or 70 units, or even probably 50 units. These students do not do as much as they should when their ability to work is considered. We might very well suppose that three students have 100 units of potential power each, and one of the three turns in 98 units of work; another 84; and the third 70 units. The second and third are not then doing the amount or quality of work we can fairly expect in the light of their potential capacity. They are not doing mental work effectively.

In our high schools we are rapidly getting our requirements for individual students on a scientific basis. Pupils with approximately equal mental ability must do approximate work. The same quantity or quality of work to be produced by three pupils, say 80 units in each case, is not estimated a "B" for all students irrespective of their difference of ability. A student with 120 units of ability who hands in 80 units production may be failed, whereas 80 units of production by a student of 100 units of capacity would be judged as creditable work, and one of 80 units production with an 80-unit mind should receive definite commendation for his work.[1]

The college instructor is a genuine teacher to the extent that the gap between potential power and actual accomplishment in each one of his pupils is small and tends to grow smaller. That wide gaps of this kind exist in most students in the average college of liberal arts could probably be neatly demonstrated if time and money for such a research were available.

---

[1] Of course we do not mean literally that intellect need or even should be thought of as a collection of units. This phrase is used only to make plain that differences in intellectual ability, and the relation of these differences in college work is of importance. Neither can we describe college work in terms of units. It would be a clear gain if we could. The point is that the relative efficiency of students' work, differences of mental ability counted in, is better than considerations of work difference of mental ability neglected.

WHY THE GAP EXISTS

The size of this difference between what a student could do and what he does do is a function of many factors, among which are undoubtedly, (a) the irrelevance of the subject matter to the purposes of the student, (b) the successful competition of many extra-curricular activities, (c) the inadequate manipulation of the laws of learning by the teacher and the like.

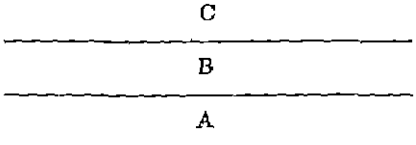
Now the laws of learning or the laws of correct teaching operate as truly in the college classroom as they do in the primary school or the eighth grade. Lack of sound pedagogy in the classrooms of a university can not be defended on any ground. In fact there is a growing realization on the part of many that skill in teaching as teaching must be highly prized and honored by college administrators.

Instruction on "How to Teach" is being provided for the younger members of several university staffs already.

DATA ON ATTENTION

One of the necessary conditions of learning is attention of a high level on the part of the learners. It may be properly asked, "What are the attention habits of college students?"

The following data may be interpreted as very strongly suggesting that one reason for the gap between what could be done and what is done by college students is that the kind of attention (or lack of it) given by students in the college classroom hinders if it does not make impossible, rapid or accurate learning. If in this simple chart, line A means a very low degree of attention; line B means the kind of listless, passive attention that one gives to familiar surroundings while riding; line C represents a rapt attention, such as one gives to the reading of a telegram or to the tense crises in a foot-ball game we can see that the amount of time spent in a recitation or lecture with attention of A type or B type is really but little more than time out. The kind of learning required in college work takes place rarely if ever except when the matter in hand is attacked with type C attention. What per cent of a 50-minute recitation is spent by a student giving



type C attention? What per cent type B or type A? No one knows, and of course the individual variations are very wide indeed. A student does not know with accuracy the amount of time spent in the subject matter with these types of attention. His opinion, though, is probably better than a guess. On the whole he is apt to report the matter with errors favorable rather than unfavorable for himself.[2]

The writers selected groups of college juniors for study. The students were on the whole among the "better sort." Their opinions concerning the kind of attention actually paid in classroom work is reported in the following tables:

TABLE I

DISTRIBUTION OF KINDS OF ATTENTION IN PER CENT OF TOTAL RECITATION TIME FOR CULTURAL COURSES, COLLEGE OF LIBERAL ARTS

%	A	B	C
5	5		5
10	9	1	10
15	8		12
20	11	2	9
25	3		6
30	1	1	5
35	7	2	3
40	2	1	1
45	1	4	
50	4		
55	1	12	
60	1	2	
65		9	
70		4	
75		5	
80		4	1
85		3	
90		1	1

A, B, and C refer to kind of attention as defined in context; % refers to % of average recitation. To be read five students thought 5% of the recitation time involved type A attention; nine students thought 10% of the time involved type A attention, etc.

These distributions give a spread of the type of attention juniors give in the average recitation. The central tendencies are: (1) that high, effective attention, type C, is given 15%-20% of the time; (2) that a lower, less acute attention is given 55%-60% of the time; (3) that between 15%-20% of the time practically no attention is given at all. Little learning can go on when attention of type A or B is given. Obviously a central improvement in college instruction would be to make those changes in class procedure which would tend to cause a large percentage of recitation time to involve that type of attention which is essential to the learning process.

---

[2] For data on the unconscious tendency of over-estimation see: Pitfalls and Rating Schemes, Knight-Franzen (Journal of Educational Research, May, 1922.)

When we have a group of students in the college of commerce estimate the percentage of time in the recitation given to the three types of attention, we get a slightly different result. The central tendency of the lowest kind of attention drops 5%. The middle 50-50 type of attention takes up about 50% of the time. The effective type of attention is the rule about 25% of the time, slightly higher than in orothodox cultural courses. These data are spread in Table II.

TABLE II[3]

College of Commerce

%	A	B	C
5	8		
10	3		3
15	3	3	2
20	11	1	1
25	2	2	9
30			4
35		1	1
40	1	1	2
45		1	
50		10	2
55		2	
60		3	1
65		1	2
70		2	1
80			
85			
90			
95			

When, however, we go over into the definitely vocational courses of law, medicine, and engineering, the opinion of the students as to the kind of attention paid in class is in startling variance with comparable opinion of students in the college of liberal arts. Data on 50 law students is spread on Table III.

---

[3] Read the same as Table I.

TABLE III

LAW COLLEGE

%	A	B	C
5	21	15	
10	14	8	
15	6	6	
20	4	3	
25	1	3	
30	0	2	
35	1	0	1
40		1	0
45		0	1
50		0	2
55		2	0
60			2
65			3
70			4
75			8
80			15
85			3
90			1
95			

Summary of these data: The drift toward effective attention habits is significantly paralleled by drift in curricula toward vocational content. This means, in the opinion of the writers, nothing more or less than conscious motivation causes large fluctuations in mental production.

A SECOND INVESTIGATION

The above data are worthy of consideration despite the absence of objective measures of mental production. We now will describe an experiment in which adequate objective data further support our contention.

THE SUBJECTS OF THE EXPERIMENT

The subjects of this experiment were ten college freshmen who had been subjected to a strenuous regime of humiliation and fatigue duty during their probation period as pledges of a fraternity.[1] A brief summary of this regimen is in order here.

The first day of probation each one of the ten subjects was given a rather strong dose of physic (two Hinkle's pills). They were not permitted to shave or bathe during the entire period of probation. A fake branding, very realistically carried through, which, according to the testimony of the chairman of the initiation committee was obviously a considerable mental shock, was a part of the program. Each freshman had to carry with him wherever he went some rather weighty or bulky article as a brick, a pail, a piece of gas-pipe, etc.[1] Besides these quantitatively rather unmeasurable indignities, which further included generous doses of "paddling" with barrel-staves, the eating of raw liver (ostensibly dog-meat) sprinkled with asafoetida; the application of embarrassing, rather humiliating names to the pledges, etc., perhaps the more significant features from the point of view of this experiment are that the men were permitted to sleep only from one to two hours out of every twenty-four, and that during each of their twenty-two waking hours they were compelled to do menial or nonsensical tasks about the house, or more frequently to go for long hikes of 12 to 15 miles. In addition to all this, they were required to attend to their regular university work.

The probation period began on Monday, and the experiment to be described here was carried out on the following Friday late at night. Here we have a group of workers, fatigued, harassed, and working at a time when the social customs of a college campus make work unusually distasteful. Eagerness for entrance to the fraternity was ostensibly related to scores made during this work period. Thus the cards were stacked against this group except for motivation.

A few days later a group of 54 college juniors (mostly women) were given a similar but shorter test on the same material as a means of securing comparable data from "normal" individuals.

Comparison of scores made on intelligence tests showed no significant differences between the two groups. The juniors were working under good conditions but with no special motivation. Thus the cards were stacked in their favor for no motivation other than those operating in a normal college recitation.

MATERIAL AND PROCEDURE OF THE INVESTIGATION

The materials used were the Thorndike addition sheets, each of which has forty-eight columns of ten addends with 1's and 0's omitted. "Any successive five of the columns are of a difficulty nearly, if not exactly equal."[1] These sheets were arranged in pads. There was no possibility of the subject's remembering the answers, for there was no possibility of identifying the seven different kinds of sheets.

Our handicapped but motivated group was told by a member of the fraternity that they were to assist in a scientific experiment and that they were to push themselves to the limit as scores on the tests were involved in final election. Desire to qualify for the fraternity was the motivation. The attitude of the group as a whole was one of cordial co-operation. At the signal "Go!" they were to begin adding and to continue until a signal "Mark!" was given. This was done at the end of every five minute period.

---

[1] See Garth, T.R., Mental Fatigue, Archives of Psych. No. 41, August, 1918.

Twenty-four such five-minute work periods were given with a ten-minute intermission at the end of the first twelve periods, at which time the subjects were given light refreshments. The "Mark" signal was for the purpose of getting an introspective check on the feeling-tone of the subjects as well as to measure progress by units of time. They were told to mark A, B, C, D, E, or F, where A signified a feeling of general well-being, no negative feeling-tone or discomfort; E was to be given the significance of intense aversion or high negative feeling-tone; C was to mean a rather neutral, don't-care-one-way-or-another attitude. Durng the last three periods a strong effort was made to obtain an end spurt by urging the men to put all they had in them into the work.

AVERAGE AND INDIVIDUAL SCORES OF THE TEN TIRED FRESHMEN ADEQUATELY MOTIVATED

The raw data give a composite work curve that is very nearly a straight line (See Figure 1) for both the total amount done and the amount correct. Nor is there any noteworthy falling off in accuracy. The data from which this curve is constructed are shown in Table I.[5] The last five-minute period does show a smaller per cent correct than do any of the other periods, but the difference is slight. The curve of the percentage correct is shown in Figure 1 at the top. The straight horizontal line through the curve represents the arithmetic mean of the percentile values for each five-minute period. Whatever decrease in total amount or amount correct occurred may be shown by comparing the work of the first 12 periods with that of the last 12. Table II gives the data for each person so compared as well as the totals for the entire 24 periods. The percentage of correct work for the first 12 periods is 85, that for the last 12, 83—a decrease of but two per cent!

The individual work curves are shown in Fig. 2. Perhaps their most striking feature is the consistent downward trend in the curve of satisfyingness and the lack of a significant corresponding downward trend in the output. The curve of satisfyingness[6]—in the case of every individual except Fitz—shows a marked decline with the progress of the experiment. In his case the intermission for refreshments proved to be annoying rather than satisfying.

---

[5] As appears from the individual work curve (Fig. 2) subject D—came four periods late. In order to calculate averages for the whole group, his total output was weighted with his average production per five-minute periods for the twenty periods that he actually worked. This procedure would seem to be fair. The general conclusions are not in any case different than they would have been without such weighting.

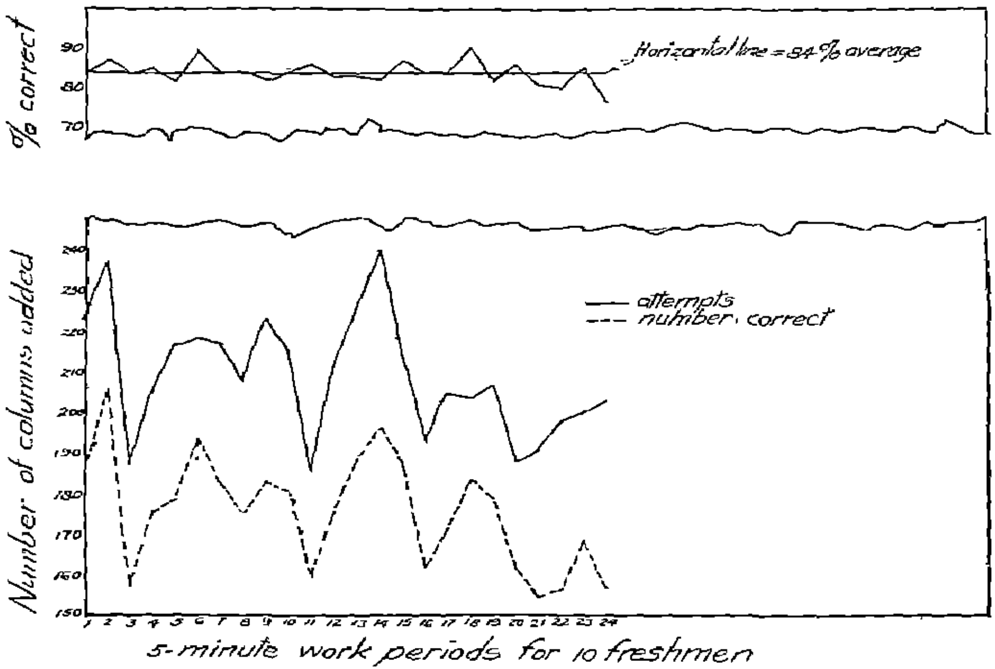


Fig. 2
Total mental product of 10 freshmen in terms of attempts and number and % correct.

TABLE I

Showing the amount of work, the amount correct, the amount incorrect, and the per cent correct for the ten freshmen in terms of columns added per five-minute period.

Period	Total Am't.	Total Correct	Total Incorrect	Per cent Correct
1	223	188	35	84
2	237	206	31	87
3	188	158	30	84
4	206	176	30	85
5	217	179	38	82
6	219	194	25	89
7	218	183	35	84
8	208	175	33	84
9	223	183	40	82
10	216	181	35	84
11	186	160	26	86
12	212	176	36	83
13	227	189	38	83
14	240	197	43	82
15	215	188	27	87
16	194	162	32	84
17	205	172	33	84
18	204	184	20	90
19	207	179	28	82
20	188	162	26	86
21	191	155	36	81
22	198	157	41	80
23	200	169	31	85
24	203	157	46	77
Av. per person	21.9	17.8	4.1	——
				Av. 84

TABLE II

Showing data for 10 individuals in terms of number of columns added.

Subject	Total Amt.	Total Correct	Per Cent Correct	Total Amt. 1st Half	Total Correct 1st Half	Total Amt. 2d Half	Total Correct 2d Half	Marks Indicating Feeling-tone					
								A	B	C	D	E	F
Fitz	536	440	82.09	257	205	279	235	23	1				
Fo	490	429	87.55	247	216	243	213	5	8	3	4	3	1
Gi	566	503	88.86	329	292	237	211	7	5	4	2	3	3
Har	536	459	85.63	277	233	259	226	9	3	5	7		
Bai	352	276	78.41	182	135	170	141	13	5	3	2	1	
An	581	478	82.27	282	237	299	241	13	3	2	3	3	
Gr	518	466	89.96	264	236	254	230	7	5	11	1		
Ne	705	571	80.99	358	300	347	271	4	14	6			
Bo	312	229	73.40	150	128	162	101	2	2	5	3		8
DeS	464	417	89.87	220	200	244	217	1	6	7	2	6	3
Av.	506.0	426.8	84.	256.6	218.2	249.4	208.8	70	51	46	24	16	15

The curves show no wide divergences of amount correct from total amount done—they practically parallel each other from one five-minute period to the next throughout the experiment.

The features of the work curve contended for by Kraepelin and his followers are not in evidence—at least not consistently for all the individuals. There is little if any more fluctuation upward at the beginning or at the end to indicate "Anlauf" or "Schlussantrieb" than chance would account for. Subjects Fits—, Fo—, Des—, Bai—, An— and Gr— show an increase both in amount done and amount correct from the first to the second period. Bo—, Ne— and Hav— on the other hand show a decrease, while Gi— does not change in total amount done. The terminal spurt is no more in evidence, notwithstanding an effort to produce one by encouraging and urging the men to increase their speed if possible. The amount done shows an upward trend for Fits—, Gi—,DeS—; Ne—, Hav—, Bai—, Gr— and Bo— however, exhibit a downward curve. Nor did the intermission cause any uniform change in output.

Fig. 3 shows graphically in another way the relation between output and feeling tone. The curves were constructed by dividing the total number of A's, B's, C's and so forth, into the number

---

[8] The range of feeling tone indicated by each subject is plotted at the right of the graph, as indicated by the letters A, B, C, D, E, F.

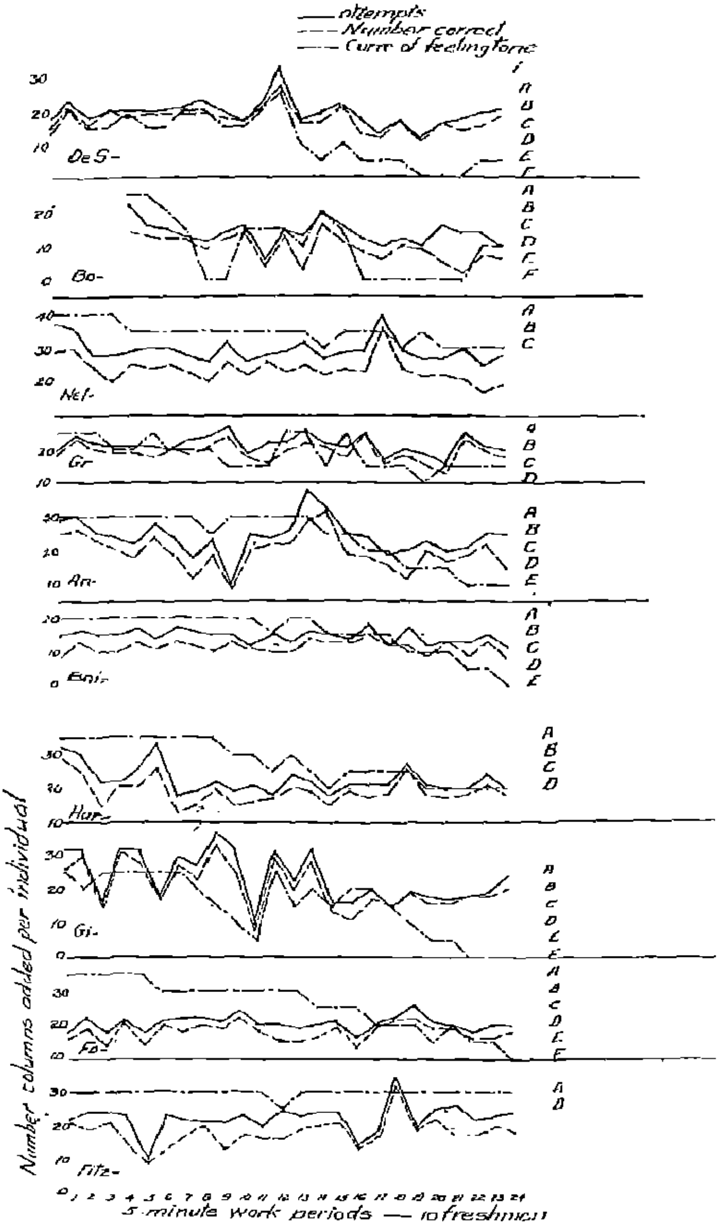


Fig II
Comparison of attempts, number of columns correct and feeling tone

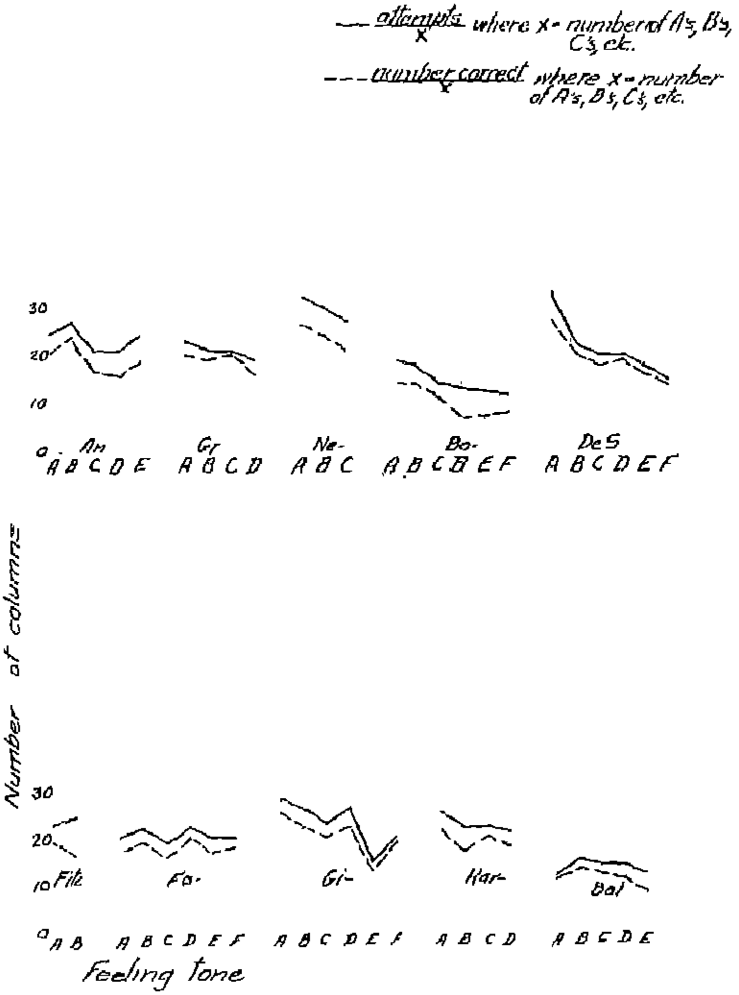


Fig. 3
Comparison of output and feeling-tone, for 10 freshmen.

of columns marked by each, and plotting these results against the letters on the abscissae. If feeling-tone correlated with output there should then be a consistent more or less parallel downward trend for the two curves. Some of the curves—those for Bo—, DeS—, Hav—, Gr—, and Ne—, do show somewhat of a downward trend. Fo— and Bai— show practically no change; Gi—, however, increased markedly in output when his feeling-tone was most negative; this is also true to a somewhat less extent for An—.

SUMMARY

The experiment was carried on for 24 five-minute periods—that is, two working hours. The ten subjects had a total output of 5060 columns added, of which 4268 or approximately 84 per cent were correct. The average total amount added was approximately 21 columns per five-minute period per person; the average amount correct per five-minute period per person was approximately 18 columns. The range of individual achievement is from 40 columns attempted with 36 columns correct for subject Nel—, to ten attempts and two columns correct for subject Bo—.

CONCLUSIONS[7]

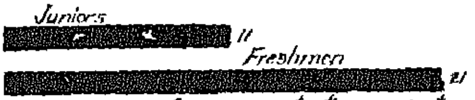
1. A two-hour work period, with this kind of material for individuals who are already physically tired, is not sufficient to decrease the output to any appreciable extent.
2. There is little or no apparent relation between feeling-tone and amount of work done.
3. There is no evidence of a warming-up period or of initial spurt.
4. Accuracy is little affected.
5. In spite of an effort to obtain an end spurt, none was obtained.

DATA OF FIFTY-FOUR RESTED JUNIORS WITH ONLY CLASS-ROOM MOTIVATION

The essential facts for this group are shown in Table III. The material used for the group was exactly like that used for the ten freshmen. Fig. 4 shows graphically the work curves for the group, both as to total amount and amount correct for each five-minute work period. There is little that is noteworthy concerning these data except the very greatly reduced output as compared with the freshmen. Possibly this is to be accounted for on the basis of class-room 'set' rather than on the ground of less ability in adding. There is nothing to warrant the assumption that the freshmen as a group were more skillful at this sort of thing by reason of previous training. The differential of motivation and mental attitudes is the explanation of the differences of an average of 21

---

[7] These data give insight into certain aspects of "Mental Fatigue." We are using them only to show amounts of work as related to motivation.



Comparison of average output per 5-minute period for juniors and freshmen

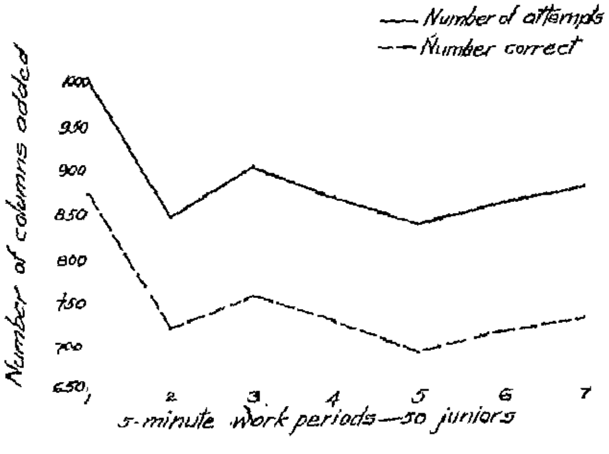


Fig 4

Total mental product of 50 juniors in terms of attempts and number correct.

columns per five-minute period for the freshmen as against an average of 11 for the juniors.[8]

The percentage correct for the juniors is in general the same as for the freshmen. It will be observed, however, that the decrease in amount correct is approximately the same for the juniors as for the freshmen. There is no reason to believe that this decrease is due to fatigue; the test was given at 8:00 a. m., the first hour of the academic day.

TABLE III

Data obtained from 64 college juniors in terms of number of columns added:

No. of Periods	Total Amount	Total Correct	Per Cent Correct	A.M. of Total Amt. Per 5-min. period Per Person	σ Of Total Amt. Per Person	A.M. of Amt. Correct Per 5-min. period Per Person	σ Of Amt. Correct Per Person.
1st 5 min.	998	872	87	18	4.18	16	4.67
2d 5 min.	843	720	84	16	4.90	13	5.25
3rd 5 min.	902	765	85	17	5.46	14	6.14
4th 5 min.	868	725	82	16	4.92	13	5.05
5th 5 min.	836	690	83	15	4.96	13	5.20
6th 5 min.	857	710	83	16	4.71	13	4.96
7th 5 min.	876	726	81	16	5.00	13	5.53
Av.	6178	5198	83.125	16.3		13.6	

CONCLUSION

The above data support the following assertion: The difference between a genuine motivation such as the freshmen had, and the kind of motivation a college class contains, is a difference which not only offsets extreme fatigue, but further offsets freshman *vs.* junior ability, and in addition produces twice as much work per unit of time with equal accuracy.

---

[8] The juniors were just an ordinary teaching group, a typical learning situation.

INFLUENCE OF MENTAL LEVEL IN THE FORMATION OF BOYS' GANGS[1]

By M. LaVinia Warner,
Ohio University

It has been noticed that certain boys group together; that a boy will have certain friends and not seem to have any other very close friends. There may be several circles of friends among the boys in a community and the circles will not intersect.

This seems to indicate that there is a fundamental cause why certain groups of boys are found together and that it is not chance but some underlying law that groups boys. The boys are brought together by churches, schools, and the like, but the churches and the schools do not classify them into the different groups and gangs. Hence the question arises: Why are certain boys found in one gang and not found in some other gang?

Dr. Charles B. Davenport in speaking of the mating of the feeble-minded says, "The meager social life, the customs of their parents, the natural ostracisms of the higher classes, and the individual's preference for a congenial mate induces in marriage among the mentally deficient." (2)[2] He found that the more frequent results of immigration of a feebleminded is his marriage into another defective strain in another part of the country. Altho in different localities, their social standards have been practically the same, for social standards are determined by mentalities as well as by experience, education and wealth. Consequently their social standards have been influenced mostly by mental level and inmarriages of the defective strains have resulted. "The social laws and natural preference of the individual is so powerful that like will consort with like even in outmarriage." (2)

Dr. Davenport says, "The law that like tends to marry like interferes with the beneficient tendency (that the introduction of new blood would lead to better progeny) of outmating. The consorts selected from outside are frequently quite as defective as those who select them." (3)

Miss Mina A. Sessions gives the same ideas of marriage selection as those given by Dr. Davenport. Miss Sessions says, "The reason for so many cousin marriages are probably twofold. First in their own stock they find the most congenial companions, and second, they are largely forced to marry each other for the simple reason that no one else will marry them or have anything to do with them. When they do marry outside the family, they marry some member of a weak strain." (10) These investigations lead us to believe that "Birds of a feather flock together," even in marriages of the feeble-minded.

---

[1] This article is a condensed form. Complete article may be found at Carnegie Library, Ohio University, Athens, Ohio, and in Columbia University Library, New York City.

[2] Numbers throughout this thesis refer to Bibliography listed at the end.

The same idea holds true in marriages of the deaf. Dr. Fay says, "In no human relation is the truth of the proverbs, 'Birds of a feather flock together' and 'A fellow-feeling makes one wonderous wise!' more fully exemplified than in the friendships and marriages of the deaf."

In his study of the divorces of the deaf, Dr. Fay found the proportion where both partners were deaf to be 2.5 per cent and where one was deaf the other hearing, the proportion reported is 6.4 per cent. This shows that the matings were more successful where both partners were deaf than in cases where one was a hearing person. The deaf partners were more congenial than one of them could have been if he were connected by hearing and speaking with those around him. (5)

The deaf have a common mode of expression. This enables them to appreciate and enjoy one another. Each is more or less barred from the hearing folk and he is glad to have a companion similar to himself. Their deafness causes them to have common experiences. It is the effect of their deafness that attracts and holds them in companionship.

Dr. Davenport and Miss Sessions determined for the feeble-minded and Dr. Fay determined for the deaf, that there are common factors at work in bringing and holding together each of these respective types of people. This suggests that the same principle could be applied to other types.

In order to make a study of the factors determining boys' gangs, various groups of boys who have been brought into court and who were known to be close companions by court officials, social workers and school officers, have been investigated. The data was derived from school reports, court records, and the records filed by the Children's Service Bureau, Youngstown, Ohio. The social histories are the reports of investigations made by the bureau's special social workers. The pyschological examinations were made by Dr. H. H. Young, director of the Children's Service Bureau and his assistants in psychology.

Thirteen groups and seven pairs of boys were tabulated and sixty-six different individuals were included in the study.

The ages in each group were computed to the date of crime committed by that group, or if no particular date could be obtained, a given date was taken. So in each group both the chronological and the mental ages of each child are given for the same date as that of his companion.

REPRESENTATIVE CLUBS AND GROUPS

THE FRANKLIN CLUB

A group of boys traveling about the city doing many little things that were not to be favored, attracted the attention of court officers and social workers. These boys were well equipped for various sorts of business that require no arrangements with the other party concerned.

The gang had three horses and an old wagon. They drove about the city picking up bottles, rubber and the like and "stealing oranges, bananas, whips, pop, beer, whiskey, cigarettes, copper," etc.

When about nine years old, A. S. one member of the gang, told of cutting the windows at a gun shop and how they were chased by a policeman. He said that they were caught twice and that when caught they gave wrong names. He said that they generally drove the best horse so that they could make a better get-away. They got up at two or three o'clock in the morning and "went out picking," returning about five o'clock.

A. S. gave many instances which show that they are a real gang. Their environment leads them to do anti-social things and with their low mentalities, they do not always make a successful get-away. In all probability they are caught many more times than they would be if they were of higher mentalities.

Their main work seems to have been stealing automobile parts from one party and selling them to another. In one case they would take the materials back and sell them to the "Jew" from whom they had stolen them.

The gang was first noticed in particular when they broke into a jewelry store window in order to steal some nickels that were on display in the window. The boys got the idea of cutting the glass from a picture show. The actor cut the glass with a glass cutter, then tapped it lightly until a smooth piece fell out. The boys had gotten some glass cutters from a twenty-five cent store the day before. They had stolen some of the cutters and had bought some. They were not as successful as the actor on the screen, however, in cutting the glass. A policeman came along, arrested them and took them to the Detention Home. They were given a hearing in the Juvenile Court and the judge suspended sentence in each case.

A few days later a boys' worker formed this group of boys into a club, later called the Franklin Club. The boys' worker tried to interest them in athletics and games of various kinds. They were taken to the park on hikes and in every case their individual interests were studied and appealed to as far as possible.

For a period of a few weeks after these boys had been organized into the club, they seemed to be doing well, but they were soon back to some of their former habits. About a month after their organization, some of them played truant and some one stole from a store in the community. Comparing their record of the past year with their behavior at this time, little improvement was found. After working with these boys and not getting the results which the boys' worker hoped to get, he took them to the Children's Service Bureau for mental examinations.

The following table shows the age, mental age and I. Q. of each at the time of breaking the jewelry store window:

TABLE I

THE FRANKLIN CLUB

	Chronological Age		Mental Age		
Name	Years	Months	Years	Months	I. Q.
J. C.	10	0	7	10	78
J. F.	11	9	7	8	65
R. L.	13	2	8	1	61
F. P.	11	5	7	11	69
A. S.	11	0	7	2	65
S. V.	13	1	9	1	69

Table II gives the mean variation of the chronological ages to be eleven months and that of the mental ages to be only five months.

TABLE II

AVERAGE AND MEAN VARIATION

	Chronological Age		Mental Age		
	Years	Months	Years	Months	I. Q.
Average	11	9	7	11	68
Mean Variation	0	11	0	5	4.16

Table I shows that the boy with the highest I. Q. is the youngest chronologically and that the boy with the lowest I. Q. is the oldest chronologically. The case of S. V. who ranks a year and more above the others in mental age, is explained in his record which shows that he is slow and easily led by others. Dr. Young states that S. V. is incapable of making his own decision. He is highly suggestible and is largely dependent upon his immediate environment and associates. His easy way and lack of will power is probably the cause of his being led by lower mentalities. Taking the group as a whole, the tables show that the mental ages bear a close sameness.

There are other gangs of boys in the neighborhood, but the Franklin boys are not found with them. The Franklin boys attend regular public schools, and this chance to meet other companions does not seem to affect their choice. Altho they come in contact with the higher boys, the duration of their companionship is so short and non-social that no noticeable event takes place.

No doubt, the mentally higher boys commit little crimes, but they bar from their company any one who is liable to expose them. Consequently the type of mind found in the Franklin club has no place with the higher type child. Then the lower level boy would not be able to contribute to the success of the gang and being a misfit he would go where he could be nearer the top in action and where he would be on an equal with his fellows.

The public school attendance department shows that in cases of truancy among the Franklin club, the boys were either alone or in company with another member of the club. However, the record of J. F. shows that he played truant three times with two other boys but their mental ages do not differ from J. F.'s more than six months and the I. Q.'s are within the range of seven units.

Court records show that boys from this club have committed ten different crimes in company with one another, but there is no account of any one of them being arrested in company with a boy who rates with normal mentality. Three boys from the Franklin club, however, have been in court with boys from other groups which are included as subjects for this theme and in each case the other boys rate below an I. Q. of 69 and with a mental age not differing a year and seven months from that of the Franklin boys, and within six months on the average.

GROUPS NUMBERED SEVEN, ELEVEN AND THIRTEEN

Of the boys examined in group seven, the one with the highest I. Q. is the youngest chronologically, and has the highest mental age. Altho the difference in mental age is greater than that in the Franklin club, yet the same tendency that like mental ages tend to group together is found here.

TABLE III

GROUP SEVEN

	Chronological Age		Mental Age		
Name	Years	Months	Years	Months	I. Q.
C. B.	15	0	10	8	71
J. B.	15	8	8	6	54
J. A.	15	5	No mental rating		
J. B. A.	13	3	No mental rating		
R. C.	15	6	9	4	60

C. B. and J. B. are very close friends, C. B. being the follower and admirer of J. B. J. B. is so low mentally that he will jump into any undertaking without fear of consequences and C. B. is low enough to follow the example. C. B. is not troublesomely inclined and he would probably cause society less trouble if it were not for his close companionship with J. B. Although C. B. is not so low mentally as J. B. yet because J. B. is older and larger and is able to do striking things, C. B. easily falls in line with all his misdemeanors.

Dr. Young says, "Time and again through the interview concerning C. B.'s past history, he referred to how much he thinks about J. B. taking money. He seems to be much impressed and to think much about how J. B. takes money from his father. The repeated reference to this together with the outline of what and how C. B. himself has stolen and what induced him to steal, seems to indicate that he has a dominating mental complex about J. B. and his stealing which with his low mentality is very largely, if not entirely the cause of his own stealing."

The individual records of C. B. and J. B. are very similar. Their histories of theft and night wanderings are practically the same. It is not exactly an influence which J. B. has over C. B., but it is a kind of unconsciousness on J. B.'s part and a thoughtless following on the part of C. B.

J. A. and R. C. have also led C. B. to steal. J. A. was not given a mental examination, but his school record shows him to be retarded more than three years and he has a long court record and a boys' industrial school record. His father reports that he cannot control him.

J. B. A. had no mental rating either, but his school record shows poor attendance and the principal reports that J. B. A. is in one day and out for weeks, and that all his school work is a complete failure. He is pushed on in the grades and does not fit anywhere. He is three years retarded in all his subjects.

The histories of J. A. and J. B. A. and the mental ratings of the other boys show the group to be below normal and their mentalities to be somewhat equal. As in the Franklin club we have no account of these boys chumming with mentally higher boys or boys of the normal type.

J. B. and C. B. appear in Group Thirteen with A. S. a member of the Franklin club and with two other boys whose mentalities are undetermined, but whose social standards are rated by social workers with those of their companions. Taking these boys as a group, we have a difference in mental ages of two years and six months and a difference in chronological ages of three years and five months. All are so mentally equipped that they are able to enjoy the same things. Each is in a group where like types are found. The I. Q.'s of these boys are below 71 and the I. Q.'s of their friends are below 78.

In the study of Group Eleven, it was found that the chronological ages are in an inverse ratio to the mental ages.

TABLE IV

GROUP ELEVEN

	Choronological Age		Mental Age		
Name	Years	Months	Years	Months	I. Q.
A. K.	15	1	8	10	59
S. J.	13	8	9	7	70
G. B.	12	7	10	8	85

The experience given the older children by living longer in the world tends to place them on an equal basis with the younger children who have higher mental ages. Consequently the experience on the one side and the experience and the higher mental ability on the other enable the boys to enjoy one another's company. The inverse ratio is found in Table I also, but it is not emphasized there so much. In fact, this tendency is found in all the groups.

Table V shows that with the exception of three groups and one pair, the I. Q. of the youngest boy is greater than the I. Q. of the oldest boy.

TABLE V

GROUPS OF MORE THAN THREE BOYS

Groups	Oldest	Youngest	Difference
1	61	78	17
2	92	69	23
Franklin	61	78	17
7	54	71	17
9	72	93	21
10	58	73	15
		GROUPS OF THREE	
4	62	56	6
5	80	80	0
6	56	76	17
8	92	72	20
11	59	85	26
12	58	82	24
13	54	65	11
		PAIRS	
14	69	72	3
15	59	85	26
16	82	93	11
17	62	72	10
18	62	72	10
19	105	70	35
20	63	100	37
		Average difference	15

This shows that older boys with mentalities below normal have a tendency to group with younger boys who have a mental age near their own and slightly higher. The higher mentalities of the younger boys enable them to cope with the older boys who have the experience of living longer in the world. This illustrates how mental equality exists in groups formed by natural selection and indicates that mental abilities tend to go together.

GROUP CONTAINING BOYS OF NORMAL MENTAL ABILITY

In three groups and in two pairs boys were found who rank with much higher intelligence than the other members of the groups. These conditions were explained when it was found that in each case, excepting one, the boy with the I. Q. much higher than the others was psychopathic, and he also had bad heredity which would influence his choice of society or would lower his standards of society.

In one pair, the lower type boy was twelve years and one month old and mentally eight years and six months with an I. Q. of 70 and the other was twelve years and eight months old and mentally thirteen years and four months with an I. Q. of 105. The mentally higher boy is an adopted child. Nothing is known of his father. The whereabouts of his mother is unknown. His mother left him at a Salvation Army Rescue Home when he was two years old and never returned for him. The mother is reported to be a drinking and immoral woman. The foster mother says that the boy is very bossy and always wants others to do as he wishes. His psychological record shows that "he does not have the persistance to work out a task requiring careful decision and forethought. He is psychopathic and impulsive to the extent that thru the path of least resistance he fails to use the intelligence which he has." The principal of the school where he attended reports that his work was extremely poor, his behavior was very queer and that he had "a mean disposition." His foster parents appealed to the court for a disciplinary school for the boy. They felt that he needed stricter discipline than they wanted to give him.

The family of the mentally lower boy moved into the same house with these foster parents. The boys were not acquainted before this. The younger boy is one of those nice looking, talkative chaps who is in for everything and who will undertake anything. The father does not live at home and the mother is unable to control the child. The older boy not caring to "use the intelligence which he has" would be glad to have anyone for a companion who would follow his suggestions and who would not be hampered by fear to do anything which might come up before the lads. The boys played truant together several times and once they ran off to another city together.

In another case the boy who was rated as average mental ability was considered by his mental examiners as one who would be readily influenced by his associates; that he would probably not initiate much, but might readily become a follower. He is psychopathic.

Another boy is much younger than the other boys in his group, but his mental age is about the same. He is lacking in initiative and is somewhat psychopathic in his responses. He is easily influenced and too readily submits to the wishes of others.

The boy who is not psychopathic is considered a case of adolescent instability. Possibly later he will grow more stable and will seek other boys for companions.

It would be interesting to continue the study of these boys to see if they still chum with the mentally lower boys in a few years following, but it is impossible to do that in this research.

INFLUENCE OF NATIONALITY AND PROXIMITY

A study was made of the nationalities represented in the different groups. Youngstown's greater population, about eighty per cent is foreign. In most of the groups there would be as many nationalities represented as there were cases. It was found that the nationality did not enter in. If the boys were all Italians, it was because they lived and met in the same Italian district.

Proximity and school effected the meeting of most of the groups and pairs of boys. In a few instances the boys met in the down town section where they sold papers or where they loitered around the cheap theatres. In one case the two boys lived on opposite sides of the city. They met down town several times, became very good friends and finally decided to steal enough money from their fathers to run away. In nine groups the boys of each group attended the same school, in ten groups the boys attended adjoining schools, and in one group three of the boys attended different schools found in the same quarter of the city and the other two boys in the group were the ones who lived on opposite side of the city.

This division of these particular groups would not hold true so very long, however, because the school system often transfers the pupils from one school to another and sometimes a child is placed in a Parochial school for a term or two. In many cases the children living on the dividing line of two school districts may have a better chance of getting acquainted than children attending the same school. The after school hours are not so closely supervised as the school hours and the children may have more freedom in choosing their playmates. Even on the playground at school the children are more or less held from freedom of choice by their school officers. Going to and coming from school and the free hours afterwards give the best opportunity for natural selection of companions. During these free hours the boys may wander where they like and they may play with whom they choose. If they do not like one group of boys, they may play alone or go off with one or two others who enjoy the same things which they enjoy.

VARIATIONS OF BOTH THE CHRONOLOGICAL AN THE MENTAL AGES

In Table VI we find that the mean variatio nof the mental age is less than the mean variation of the chronological age in three out of five of the largest groups, that is the groups containing more than three boys. This shows that there is a tendency for less variation among the mental ages than among the chronological ages.

TABLE VI

MEAN VARIATION AND RATIO OF MEAN VARIATION TO AVERAGE

	Mean Variation		M. V. Average	
Group	Chronological Age	Mental Age	Chronological Age	Mental Age
1	13 months	7 months	.09	.07
2	3	15	.02	.13
Franklin	11	5	.08	.05
9	17	9	.12	.08
10	16	17	.10	.14

Since the intelligence quotients are below 100, the mental ages are less than the chronological ages and are expected to be nearer together. In order to determine further that mental age has more weight than chronological age in the formation of groups of boys, the ratio of the mean variation to the average (coefficient of variation) in both the chronological and the mental ages was computed in the larger groups and a comparison was made of their variability. It was found that the mental ages were less variable than the chronological ages in all the groups excepting in Group Two and Group Ten. See Table VI. In these two groups the chronological ages were slightly less variable than the mental ages.

The average of the mean variation of all the groups and pairs was found to be 10 months for the chronological ages and 11 months for the mental ages. This includes the groups whose mean variation was higher for the mental age than for the chronological age and the pairs whose mean variations were not thought necessary to compute before because of the small number of boys included in each. The mean variation for the whole population (66) was found to be 18 months for the chronological age and 16 months for the mental age. Thus we find:

$$\begin{array}{lcc} & \text{Chronological Age:} & \text{Mental Age} \\ \hline \dfrac{\text{Average M. V. of Group}}{\text{M. V. of population}} = & \dfrac{10}{18} = .05 & \dfrac{11}{16} = .08 \end{array}$$

In some groups, the chronological ages are quite similar. This is very marked in Group Two. In this group, all of the chronological ages are in the twelfth year. The mean variation of the mental ages is greater than the chronological ages of Groups Four, Seven, Ten and Twelve.

In Group Seven, the chronological ages are very similar and the mean variation of the mental ages is greater as near as could be discovered. Two members of this group have no mental rating so the mean variation of the mental ages could not be determined in exact terms for all the members in the group.

The chronological ages of the two boys in Group Eight are nearer together than their mental ages. The chronological ages of the other boys in this group could not be found and they were not given a mental test, so they might alter this group in either direction. There is very little difference in the chronological ages in the boys of Pair Fourteen and in Pair Nineteen, also.

This shows that chronological ages might have some weight in the grouping of boys, but since there are more groups in which the mean variation of the mental ages is less than the mean variation of the chronological ages, we might conclude that there is a tendency for the mental ages to have more weight in the formation of boy's groups than chronological age.

DISCUSSION

In the Franklin Club the mean variation of the chronological ages is twice as great as that of the mental ages. In this group, mental age must have affected the grouping. Common experience enters into the grouping also. If one of the boys had lived until this time in a rural district he would not have enough experiences common to the others to fully appreciate the attitudes of the other boys altho their mental abilities are practically the same as his.

Proximity influenced the grouping in most of the cases. It even had an influence in the case of the boys living on opposite sides of the city. It was a common experience which brought them together and it was their mental ages kept them together. If one of the boys had been mentally high and the other mentally low, their common experience of selling papers down town would scarcely have been noted. Common experience is the means of bringing the boys together and mental level is the determining factor which continues to make their experiences common.

Experience, education, and social status help determine the choice of comrades among adults of the same level and their social standards may differ so that their tendencies might be in an almost opposite direction, but among children the selections are less controlled. Children have more freedom in choosing their companions because they have not lived long enough to be influenced in making their choice. Social standards, experience and education have not affected them to so great an extent.

The boys in these twenty groups were not known to associate with boys who rank as normal, altho they had the same opportunity to get the same experience. Physically, their experiences were common, or could have been, but mentally each type of mind interpreted the experience to suit its own level.

In several groups, the oldest boy has the lowest I. Q. and the youngest boy has the highest I. Q.; that is, the I. Q. varies inversely with the chronological age. This brings the mental ages to an equalized level.

Boys of the same mental level can understand and appreciate the same things. They have reached the same degree of progress in the stage of mental development. Their minds are so equipped that they can appreciate the same situations. Consequently, each is at home with the other. No one is comfortable for a very long time if he is much inferior to those about him. He must have a feeling of his own importance; that he has a place of his own in their society, or he will begin to shrink mentally away from them. On the other hand, if he can play his part on an equal basis with his companions, he will have a feeling of satisfaction and will be content to remain in their company.

A boy mentally high will not be satisfied with companions several degrees below himself. They cannot reach his ideas and he will grow tired of being tied down to their level, so he will stand aloof or seek others for companionships, except in some cases where he takes a paternalistic attitude, and assumes a sort of guardianship or big brother attitude.

A special class teacher reported that her two lowest children who were about seven years of age, would not have anything to do with one another. No doubt they were too young mentally to be social. They were mentally below four years. The older and higher type of children in the class enjoyed leading these younger boys about and caring for them. They had the paternalistic attitude and the younger children enjoyed the feeling of their protection. We enjoy being with leaders or individuals who are our superiors because we feel that we are being benefited by their presence. They cause us to live in an ideal state which we enjoy.

The low type child enjoys the company of the older children because he likes to have them do things for him. He is taking a babyish attitude.

This is not an attempt to solve the problem of group companionship in so brief a study of so few cases, but the investigation indicates that mental age is the greatest factor in the selection of one's companions and in holding groups together. It indicates that a common mental level with similar experiences as a background is a much stronger factor in group formation than any other factors investigated.

BIBLIOGRAPHY

1. Cattell, J. M. *American Men of Science*, Second Edition
2. Davenport, Chas. B. *The Hill Folk*
3. Davenport, Chas. B. *The Nan Family*
4. Dugdale, R. L. *The Jukes*
5. Fay, E. A. *Marriages of the Deaf in America*
6. Giddings, F. H. *The Principles of Sociology*
   Consciousness of Kind
7. Goddard, H. H. *The Kallikak Family*
8. Hall, G. Stanley *Adolescence*
9. Puffer, J. Adams *The Boy and His Gang*
10. Sessions, Mina A. *The Feebleminded in a Rural County of Ohio*
11. Swift, E. J. *Youth and the Race*
12. Thorndike, E. L. *Measurements of Twins*
13. Winship, A. E. *Johnathan Edwards*

LEGIBILITY OF BODONI, BASKERVILLE ROMAN, AND CHELTENHAM TYPE FACES

By Harold E. Burtt and Coryne Basch
Ohio State University

INTRODUCTION

The importance of type legibility for advertising or general reading needs no comment. The outstanding variables conditioning legibility are size, case and face. The present study is concerned primarily with the last of these and is confined to an intensive investigation of three common type faces.

There are five methods (as far as the writers are aware), for measuring legibility: maximum distance at which type can be read, time taken to read a passage, number of letters read in a tachistoscope or minimum exposure at which they can be read, minimum illumination under which type can be seen and extent to which letters can be thrown out of focus and still be identified. The first two of these methods have been the ones most frequently used. The present study used the last of the five methods mentioned.

The most extensive study of type faces is that of Roethlein.[1]

She presented a page of 28 letters at a distance beyond the limits of legibility and then advanced it toward the observer by 20 cm. increments, recording at each stage the letters that could be identified. This method made it possible to accumulate a large amount of data in a reasonable time and a considerable number of different faces were studied. Cheltenham, however, was the only one which she used that was involved in the present study so no comparison is possible.

Roethlein found that legibility depended on form, size and heaviness of the letters and on the margin, position in group and on the shape and size of adjacent letters.

Scott[2] found that a longer time was taken to read a light-faced type than to read a heavy-faced type.

Adams[3] performed an experiment on speed of reading a series of advertisements set in a wide variety of type arrangements. The material was too complicated to throw much light on the specific problem of type face.

The present study tried to abstract from all variables except type face. Individual letters were used, all of the same size.

---

[1] Roethlein, B. W. *The Relative Legibility of Different Faces of Printing Types.* Amer. J. of Psych. 1912, 23, 1-36.

[2] Cf. Hollingworth, H. L. *Advertising and Selling.* 1920, p. 73.

[3] Adams, H. F. *Advertising and its Mental Laws,* 1921, p. 214.

MATERIAL

Three type faces were used in the experiment—Bodoni, Baskerville Roman, and Cheltenham Medium. These were used in both upper and lower cases. All the letters were fourteen point so that the size variable was constant except in so far as the upper case letters often occupied greater area than the corresponding lower case of the same point. The letters were cut from sample booklets of the American Type Founders Company.

Each letter was centered individually on white unglazed cardboard 6x6 inches for exposure. It was impossible to get a complete alphabet in all the faces and cases. Consequently, the results considered below were based on all the letters of the alphabet with the exception of J, K, L, Q, U, X, Y, Z. Several of these letters are rather infrequently used anyway. The remaining eighteen letters in the three faces and two cases make one hundred and eight letters involved in the investigation. Some other letters were included but the results not considered.

The material was presented in a Focal Variator.[1] This apparatus comprises an exposure field about 6 inches square illuminated by daylight lamps inside a hood. The material exposed is projected through a stationary-convex lens which controls its size and then through two similar lenses which are arranged to move reciprocally so that the focal point of the image may be changed without altering the size. The material is projected through this differential system upon a stationary ground glass so that as the lenses move the image on the glass goes in and out of focus. The subject sitting with his forehead in a head-rest fifteen inches from the ground glass observes the images. In the present experiment a quarter-inch diaphragm was placed on the lens next the ground glass to lessen the marginal glare. This was desirable because of the small size of the letters used but did not affect the focus.

The differential lens system was controlled by turning a small crank. This crank actuated a screw 16 threads to the inch so that for each revolution an indicator moved 1-16th of an inch on a linear scale. The crank was also attached to a small wheel with its margin graduated into 100ths so that it was possible to make readings to 1-100 of 1-16 of an inch.

The data will be presented in terms of these scale readings. However, it is possible to convert such readings into terms of the position at which the image is actually in focus.

The apparatus was set during the experiment so that at zero scale reading the image was perfectly in focus on the ground glass. The callibration curve was obtained by movin gthe ground glass toward the lenses by increments of one inch and at each point determining the scale reading when the image was in perfect focus. These determinations were made by coming in both directions from illegibility until a letter showed no appreciable increase in clearness and continuing until it began to decrease in clearness. These four scale readings were averaged. This procedure was followed with five typical letters at each point and with two observers. Scale readings were plotted against distance of the focal point from the exposure field. The curve is practically a straight line whose equation by least squares is $2.09x-.64y+26.85=0$ x represents scale reading, and y represents distance of focal point from the exposure field.

---

[1] Weiss, A. P. *The Focal Variator*, Journal of Experimental Psychology. 1917, 2. 100-113.

METHOD

The experiment was conducted in a dark room. During an actual trial there was no illumination except the lights in the hood and a small heavily shaded light near the scale. Between trials there was practically no change in the illumination so that the subject was substantially adapted throughout to the ground glass field.

Prior to each trial the lenses were turned to a point previously determined where all the letters were invisible. A letter was then placed on the field at the back of the hood. At the signal "ready," the observer placed his forehead on the head rest and fixated the center of the ground glass. The experimenter then turned the crank at the rate of one revolution in four seconds, keeping time with a metronome. The observer named the letter as soon as he recognized it. If his response was correct the experimenter released the handle immediately and recorded the reading. If the response was incorrect, the experimenter said "No" and continued turning the crank until the correct response was given. The incorrect response was, of course, subsequently recorded.

The letters were given in a random order with all the cases and faces inter-mixed. Each subject worked for two sittings and in some cases three, in order to complete the entire series. The subjects were frequently asked if they were fatigued and there was considerable opportunity for rest between trials while the experimenter was recording results and changing material.

The subjects were seven students taking a laboratory course in psychology. None of them had had experimental training other than that which they had received in the course. There were five males and two females. Five of the seven had previously performed a simple experiment with a Focal Variator themselves. The vision of all was 20-20 or corrected to that by glasses.

RESULTS

The outstanding differences between the three families of letters are shown in Table I. The first column gives the observer, the second the average reading of that observer for the eighteen letters in Bodoni upper case. The next column, the readings for Bodoni lower case, and so on. For example observer number one gave an average scale reading for the eighteen letters in upper case Bodoni of 14.7 16ths of an inch. His corresponding reading for Bodoni lower case was 11.4. For Baskerville upper case 14.6 and so on. The bottom row in the table gives the average for the seven observers. The same results are shown graphically in Figure 1.

TABLE I

AVERAGE LEGIBILITY OF 18 LETTERS

Observer	Bodoni		Baskerville		Cheltenham	
	U. C.	L. C.	U. C.	L. C.	U. C.	L. C
I	14.7	11.4	14.6	12.8	17.6	13.1
II	12.9	10.9	13.6	12.0	16.4	12.9
III	12.9	10.9	16.5	10.9	17.9	13.4
IV	12.2	10.7	13.1	10.7	14.4	12.0
V	13.6	11.2	16.5	12.6	17.4	13.2
VI	12.7	11.3	15.5	12.4	17.5	13.6
VII	13.9	11.0	15.8	11.9	17.5	13.3
Average	13.3	11.1	14.9	11.9	17.0	13.1

The abscissa of the curve represents the three faces indicated while the ordinate represents the scale readings. Each observer is indicated by a different type of curve with the heavy curve giving the averages. The nine curves nearest the bottom of the chart indicate lower case and the others upper case.

The general tendency of results is obvious. Cheltenham is more legible than Baskerville. Baskerville in turn is more legible than Bodoni. The results are more marked with the upper case letters. It may be noted incidentally that throughout the upper case is more legible than the corresponding lower case. It should be borne in mind, however, that we are here dealing with individual letters whereas it is well known that entire words printed in lower case are more legible than the same words printed in upper case.

The results are presented from the standpoint of probable error in Table II.

TABLE II

RATIO OF DIFFERENCE TO PROBABLE ERROR OF DIFFERENCE

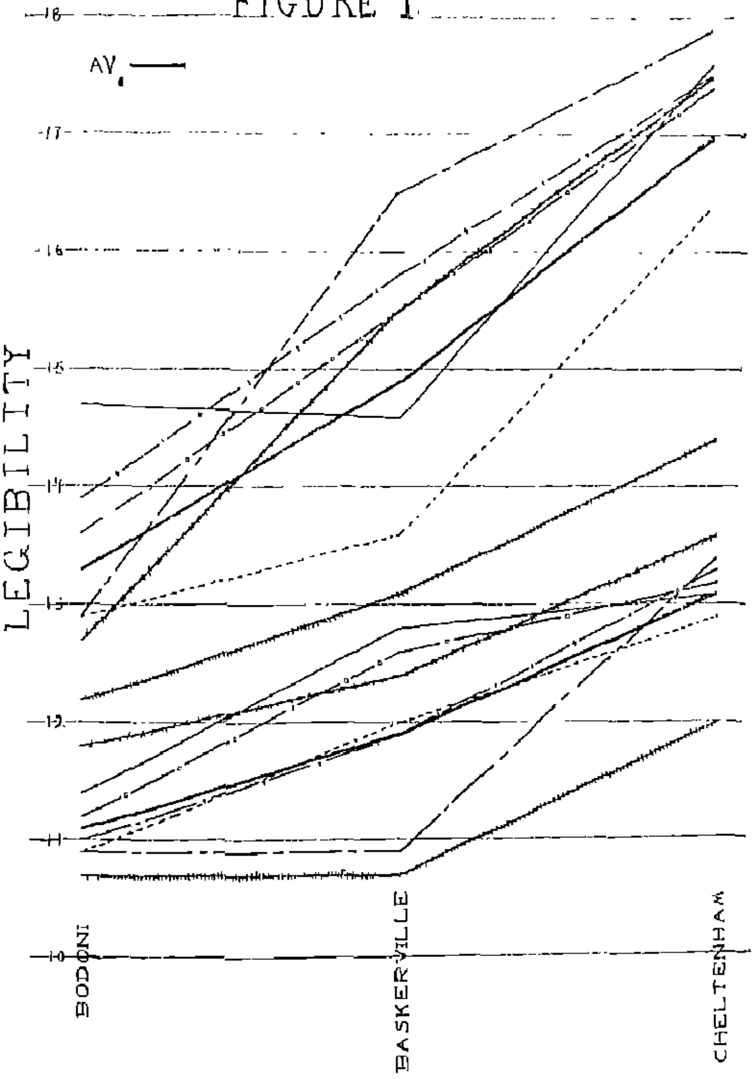
Observer	Bodoni Baskerville		Bodoni Cheltenham		Baskerville Cheltenham	
	U. C	L. C	U. C.	L. C.	U. C.	L. C
I	0.2	2.2	4.8	2.8	5.8	0.4
II	1.4	2.2	6.1	3.9	4.1	1.7
III	5.2	0	8.6	3.3	1.9	3.2
IV	2.3	0.2	4.8	3.4	2.8	3.6
V	3.4	2.7	7.6	4.1	3.2	1.3
VI	5.0	1.3	8.1	3.3	3.8	2.3
VII	4.8	1.9	8.3	5.8	4.3	2.8
Average	3.2	1.5	6.9	3.8	3.7	2.2
All cases	7.2	3.0	16.9	9.5	8.8	4.4

This table gives the ratio of difference to probable error of difference for the various comparisons and subjects indicated. For example, with subject I the difference between the average of Bodoni upper case and the average of Baskerville upper case is 0.2 the probable error of difference, while for those two faces in lower case the difference is 2.2 the probable error. The figures for the seven subjects in each column are averaged in the row marked "average." The probable errors were also computed for *all* the letters of a given face and case for all the observers, i. e., 126 (18x7) readings. The differences between the averages at the bottom of Table I may be thus evaluated and in Table II the row "all cases," gives the ratios of these differences to the probable errors of difference.

The table indicates that the difference between Bodoni and Cheltenham upper case is unquestionably significant, as the smallest individual subject's difference is 4.8 P. E. The differences between Baskerville and Cheltenham are not as marked but some of the individual differences are significant and the average difference undoubtedly so. The differences between Bodoni and Baskerville upper case are of about the same order although not quite as large. With the lower case letters the relations are much the same as with the upper case but all the differences are somewhat smaller from the standpoint of probable error. The difference between Bodoni and Cheltenham is the most marked, that between Baskerville and Cheltenham somewhat less, while that between Bodoni and Baskerville lower case is of somewhat doubtful significance.

These results seem plausible from analysis of the characteristics of the three families. Cheltenham involves relatively heavy strokes of fairly uniform width, Bodoni on the other hand, involves many light strokes and this is especially marked with upper case letters.

FIGURE I


1·8
1·7
1·6
1·5
1·4
1·3
1·2
1·1
1·0
AV.
LEGIBILITY
BODONI
BASKERVILLE
CHELTENHAM

Baskerville likewise involves light strokes though not as light as Bodoni. This corresponds to its intermediate position in legibility.

In addition to the differences between the families established on the basis of the average of 18 letters, there may be points of interest in considering individual letters in the various faces. Table III presents data from this standpoint. Each entry in the table represents the average of the seven subjects for the letter and face indicated.

TABLE III.

Letter	Bodoni U. C.	Bodoni L. C.	Baskerville U. C.	Baskerville L. C.	Cheltenham U. C.	Cheltenham L. C.
A	13.1	9.5	13.9	9.9	16.9	12.4
B	12.7	13.0	14.1	15.2	15.1	14.9
C	14.1	10.7	15.8	11.4	18.3	13.2
D	14.0	12.2	16.3	12.5	18.2	14.5
E	12.7	9.4	13.1	10.5	16.2	12.2
F	11.2	11.6	14.2	12.1	15.4	10.6
G	13.7	10.9	16.5	9.9	18.2	10.1
H	14.2	12.6	15.6	13.3	17.1	14.7
I	12.8	10.0	12.2	13.5	13.1	13.5
M	13.7	13.7	16.9	11.9	17.2	15.8
N	11.8	9.7	15.9	11.0	17.0	11.9
O	15.8	12.5	17.0	12.7	20.2	14.5
P	14.4	13.5	14.8	14.1	17.8	15.4
R	13.0	11.1	14.1	11.5	16.5	10.5
S	13.5	8.6	12.9	10.3	15.3	11.6
T	11.7	9.1	14.1	9.5	16.5	11.0
V	12.2	10.9	14.0	11.9	18.1	13.2
W	13.3	11.4	17.1	13.4	17.9	14.8
Av.	13.2	11.1	14.9	11.9	16.9	13.0

Obviously the legibility differences vary with the letters. The safest way to consider the matter is in terms of probable error. This was computed for any differences between means of the same letter in different faces that seemed of possible significance.

The differences between lower case Bodoni and Cheltenham are most pronounced with the following letters and the differences are all greater than 4.5 the probable error: S, I, V, C, A, E, W. Most of these letters involve a considerable portion of light strokes which in the Bodoni are of course very light. With lower case Baskerville and Cheltenham the ratio is 7.2 for A and 4.4 for C, with all the others less than this. With Bodoni and Baskerville lower case the ratio for I is 8.9 but the next smaller is 3.7 for P. The most noticeable fact regarding the lower case letters is perhaps the inferiority of Bodoni or Baskerville A to Cheltenham A.

This letter has three light horizontal strokes.

With the upper case the differences are relatively greater. For Bodoni vs. Cheltenham they are significant for practically all letters, but most marked in the case of V, N, B and O. With Baskerville and Cheltenham, V, E, P, S and R, all have differences greater than 4.5 P. E. With Bodoni and Baskerville the ratio for N, G and B are greater than 4.5. One noticeable thing is the poor showing of Bodoni N. It comprises two very light vertical strokes.

Another method of approach is to consider each letter's legibility relative to the others of the same face and case. This could be accomplished by ranking the values in each column of Table III. The greatest discrepancies between an upper case letter's rank in different faces are as follows: N has relatively low legibility in Bodoni compared with Baskerville. This fact has been noted above: S however ranks much higher in Bodoni than in Baskerville. V stands high in Cheltenham but low in the other two. With lower case letters the following discrepancies are large. Baskerville I is relatively high in legibility and Bodoni I relatively low. Bodoni F is relatively high and Cheltenham F, relatively low. Cheltenham and Bodoni M are the most legible in three faces while Baskerville is only average.

In general there is a somewhat closer correspondence in the relative legibility of the letters in the three faces with lower case than with upper case. Correlations were computed (rank differences squared) between the different columns of Table III. For the upper case the correlations are: Bondoni-Baskerville .47, Bodoni-Cheltenham .74, Baskerville-Cheltenham .63. For the lower case the corresponding figures are .73, .70, and .73, indicating closer correspondence in the case of the lower case letters. It would seem that upper case letters afford a better opportunity for face differences in legibility to manifest themselves.

A record was kept of all cases in which a letter was mistaken for another. These were tabulated to show any common tendencies. No error was considered of interest unless made by at least two subjects. The following confusions occurred in all three faces: lower case E read C; lower case H read B; upper case I read L; upper case R read K. The following confusions occurred in two faces, lower case; F read I, T read I; and the following in two faces, upper case: B read C; M read W; P read F. There was nothing in the analysis of results from this standpoint to clearly indicate differences in face.

SUMMARY

A study was made of the comparative legibility of Bodoni, Baskerville Roman and Cheltenham Medium type faces. The legibility of individual letters was determined separately using upper and lower case and fourteen point. The letters were presented in a focal variator and evaluated in terms of the distance they could be thrown out of focus (maintaining constant size) and still be recognized.

Greatest legibility was found on the average for Cheltenham, followed by Baskerville, with Bodoni the least legible. This was true for both upper case and lower case but the differences were more striking from the standpoint of probable error with the upper case. These differences seem plausible upon examination of the three faces, particularly with refenence to the width of strokes. Study of individual letters showed a tendency for the greatest differences between faces to occur with letters involving light strokes, inasmuch as these are particularly light in the Bodoni and Baskerville.

A COMBINATION MENTAL TEST FOR CLINICAL USE

By Grace H. Kent
Worcester State Hospital, Worcester, Mass.

I. LIMITATIONS OF STANDARD TESTS, FOR DIAGNOSTIC PURPOSES.

The *serviceability of mental tests for tentative classification of* large groups of persons has been established beyond reasonable doubt and there are many tests in common use, both individual and group, which are sufficiently accurate for such purposes as classification of army recruits, school children, or factory employees. But in the clinic no mechanically applied test should be expected to take the place of patient and painstaking individual study. Used as an instrument for such study the standard test may be a valuable aid; but it is of questionable value if we accept the findings as final and thus make it serve as a substitute for careful observation.

The concept "mental age" is an exceedingly convenient one and it is not unnatural that a servant so obliging should be expected to work over-time. Because the mental age can be so easily stated, it is almost inevitable that its significance should be exaggerated. Thus we find a tendency to speak of a child's mental age as if it were as definitely and finally determinable as his life age and there *is danger that the qualitative findings of the test—which may be* vastly more significant than any results that can be stated in figures—will be wholly overlooked.

An unavoidable source of error in the use of a norm is that it is impossible to give due recognition to the variations in the individual records from which the norm is constructed. The norm may be derived from a sufficiently large number of representative persons, but the individual variations still remain and the limits of possible normality lie close to the extremes rather than near the average or median.[1]

---

[1] I plead guilty to the failure to give due weight to this factor. In the interpretation of the data which formed the basis of the Kent-Rosanoff association test, published in 1910, I was so deeply impressed by the differences which the test indicated between normal and insane subjects that I did not fully appreciate the importance of the similarities thus indicated. *The average number of* "individual reactions" for normal subjects was seven per cent of all reactions, and the average for insane subjects was twenty-seven per cent. This is obviously a significant difference, but we cannot determine what it signifies for a given individual. The range of variation was very wide for both classes of subjects, being from zero to ninety per cent for the normal subjects and from five to ninety-seven per cent for the insane subjects. There was thus an *overlapping amounting to eighty-five* per cent between individuals of the two classes, and it is plain that no conclusions should be drawn about any individual on the basis of a test which yields such variable results.

Another source of error which should be kept in mind is the personal equation of the examiner. This factor is well eliminated in many group tests by reducing the administration of the tests and the scoring of results to a purely mechanical process. But it cannot be so easily overcome in individual tests because much depends upon the inflection with which the examiner reads a question, and upon the spirit and interest with which he presents a problem. The test may be so fatiguing to the examiner that he is unable in the latter part of the working day to do full justice either to the test or the subject. And in spite of the scoring cards and minute instructions which add more to the difficulty than to the accuracy of scoring, there remains considerable room for difference of opinion, especially in tests calling for involved verbal responses.

In order to be valid for adult subjects a test should be based in part upon faculties that develop with increasing maturity and experience as opposed to those faculties that reach their zenith during the high school period and fall into disuse after school life is completed. If we are to depend largely upon tests for diagnosis, we should have a test that would show in what way a deficient man having a "mental age" of nine years differs from a normal boy of nine. We know that the man lacks the developmental possibilities which constitute the boy's chief asset to society. We know also that the man, by reason of his longer experience and more settled habits, can do many things which the boy cannot do. We indicate the man's inferiority to the boy by assigning an intelligence quotient of fifty-six; but our criteria for determining his intellectual status are not adequate unless we can show also in what ways he is superior to the boy. Determination of the mental ability which is measured by the Binet scale, call it what we please, does not tell us what we need most to know about a clinical subject. The man of thirty who gives a nine-year mental rating may be a useful and trustworthy laborer, or he may be a serious menace to society. Two thirty-year men of nine-year mentality, equally free from delinquent tendencies, may differ very widely in their capacity for work, in their adaptability to different kinds of work, and in the amount of supervision required. The fact that they give the same mental rating is relatively unimportant since the proper disposal of their cases depends upon distinctions which are not shown by the test.

It is difficult to over-state the possible errors involved in the use of a verbal test for subjects who do not understand the language in which the test is given. It is true, under certain conditions, that a child's ability to acquire language may furnish a fair index to his intelligence, but we cannot make this assumption in a given case unless we know that the subject has had adequate opportunity to acquire the language. I recently tested many children in a mill village of which the foreign population is in the majority. The children use their own language at play, even on the school grounds, and many of them have rarely heard a word of English outside of the school room. To assign a mental age based on such showing as they can make in a verbal test would seem little short of perjury. Performance tests are useful in such cases, but more for qualitative interpretation than for mental rating.

A test stands on its own merits, regardless of norms, since the findings obtained by an unstandardized test that is well suited to the subject are more significant than a mental rating derived from an ill-adapted test. Dr. Walter E. Fernald's questions on "Practical knowledge" possess certain advantages over any standardized test known to me. Although they cover a wide range of difficulty, these questions can be asked casually and answered simply, usually in a very few words. They relate largely to matters of local interest, and the child may easily assume that the examiner is seeking information on the matter inquired about. There is no definite programme for asking the questions, as for administering a more formal test, so the examiner is free to omit any question which does not fit the case and which might make the subject feel strained and unnatural. I should like to see these questions standardized, but not unless it could be done without destroying the elasticity of the system. For although a norm is a great convenience to the examiner, it is not essential to the serviceability of the test. The age value represented by a given question or task is empirical, not absolute, and the examiner's own experience ought not to be wholly at a discount. When a test—even a standard test—is used by an experienced observer, I question the advantage of rigid adherence to rules.

No test can be considered valid except on the assumption that the subject answers each question as well as he can. But it frequently happens that a subject who is brought to the clinic for examination comes very unwillingly and with no intention of giving any information about himself. The tactful examiner can usually obtain some measure of responsiveness, but it is impossible to state the exact degree of co-operation received. The lack of co-operation cannot be estimated in quantitative terms, while its effect on the mental rating is invariably quantitative.

Among insane subjects one meets with all possible degrees of co-operation. Some patients put forth their maximal effort in the hope of proving their fitness for life in society, some offer violent resistance, and some give no response at all. There are many cases in which the qualitative findings are of considerable interest, but it is exceptional to find a patient for whom the mental rating, as such, has any significance even as a matter of record, still less for diagnosis or treatment. When the rating is high enough to exclude mental deficiency, this diagnosis is almost certain to be excluded also on other grounds. A low rating may indicate deficiency, deterioration, or poor co-operation. Extremely wide scattering of success and failure is usually associated with a psychotic condition, but it may indicate merely that some of the problems presented are interesting enough to command the co-operation of the subject while others are beneath his notice.

The examiner who administers the test usually learns something about the case but not necessarily anything that can be stated statistically. In many cases it would be grossly misleading to assign a mental age.

II. INHERENT WEAKNESS OF THE STANFORD REVISION, FOR CLINICAL USE.

No single system of tests is equally well adapted to all classes of subjects. The Binet scale is especially adapted to children of normal mentality, and in this field I can agree with Terman that "In the large majority of cases persistent silence deserves to be scored failure." School children are in general the most responsive subjects I have ever had occasion to test, not excepting psychologically trained university students, and the matter of obtaining their co-operation is not a serious problem.

But this does not hold for clinical subjects, even when they are tested in school. The mental measurement of defectives requires a much finer technique than is necessary for the classification of normal children, and at best there is serious danger of rating a child too low. The younger children who are brought to the school clinic are apt to be timid in the presence of a stranger, and the older ones are usually suspicious and sensitive. Defective speech is very common, frequently serious enough to make the child shy and self-conscious, and sometimes so marked that his answers are unintelligible. Deficient children are very susceptible to fatigue, and the examination is usually too long, except for a child rating very low, to be made properly at one sitting.

When the external conditions are favorable and when there is no lack of time, I find the lower end of the Standford scale fairly satisfactory for testing English-speaking children who are free from speech defect. If a child is too young to realize that he is being tested for mental deficiency, it is merely a question of time to overcome his diffidence and to gain his confidence. He regards the test as a school exercise, and responds to the examiner as well as to a new teacher.

Many of the difficulties encountered in the clinical use of the scale are clearly recognized by Terman and cited as exceptional situations which should be met by some special mode of procedure.[2] But in the clinic these difficult situations are the rule rather than the exception, and the means suggested for overcoming them are usually impracticable. Under ordinary clinical conditions, as when the children are brought from a distance or when they are examined in school by a visiting psychologist, it may be impossible to take the child out for a walk in order to establish a friendly feeling, to allow two sittings for an examination, or to resort to any of the measures prompted by common sense. A full "Out-Patient Day" is not conducive to accurate work. The examination is usually too hurried to be satisfactory, and after four or five equally unsatisfactory tests performed in quick succession the examiner may be too tired to take proper interest in subsequent tests.

My principal objection to the lower half of the scale when used for juvenile subjects is that it requires conditions more perfect than usually obtain in the clinic—more perfect than I have ever seen in any clinic. But for any subject able to go above the eight-year level and for an adult subject of any grade of mentality, the defects of the scale are more serious and more fundamental.

Perhaps the most delicate situation which we meet in the psychological clinic is that presented by the examination of a self-supporting young man or woman of border-zone intelligence, who requires supervisory care because of some problem relating to conduct. The subject may be painfully aware of his intellectual shortcomings, and the humiliation of being tested for mental deficiency is almost more than he can bear. If the test is to be valid, we should be able to offer something sufficiently interesting to draw his attention away from himself as a subject-who-is-suspected-of-being-feebleminded. But the tasks of the Stanford scale are strongly conducive to self-consciousness, especially the questions relating to school life. It is little less than inhuman to require a serious answer to the question "What should you do if in danger of being late to school?"

The vocabulary test furnishes a good illustration of a test that is well adapted to school children but very poorly adapted to the clinic. A bright boy may regard it as a joke when his limit is reached, and I have never seen a normal child manifest serious annoyance over the series of failures with which the test is invariably brought to a close. But these failures are disheartening to the sensitive clinical subject, and the situation becomes more tense with each succeeding question. The test is, therefore, unduly fatiguing both to subject and examiner.

---

[2] The Measurement of Intelligence, pages 124 to 135.

It is not necessary to consider each question individually.[3] The fundamental weakness of the system is that too little account is taken of the affective reaction of the subject. Terman assumes, as did Binet, that the questions can be presented in such a way as to arouse the interest of the subject. But even if so, it is a serious defect of the system, for clinical purposes, that so much depends upon the presentation. As most of the subject matter is of no intrinsic interest to subnormal adults and adolescents, the resources of the examiner are severely taxed in the effort to make the test attractive enough to hold their attention. And the rules for the administration of the test, when they are followed literally, make the examination unduly lengthy and tiresome. The examiner is expected to cover one year wholly within the range of the subject's ability and one year wholly beyond his range, thus including at least twelve questions that are frankly nondiscriminative for him, besides many more that are probably so. It is these nondiscriminative questions that are most disturbing to the subject's emotional equilibrium. He is annoyed and discouraged by those that he cannot answer and frequently feels insulted by those which are so easy as to require no thought.

It is of no small importance, in the interests of accuracy, that a clinical test should be easily and comfortably administered. At best a day's work in the clinic is usually too exhausting to be done as it ought to be done. One who is engaged in collecting material for a new test of his own devising may well think nothing of working eight or even ten hours at high tension, because all thought of fatigue is overshadowed by the exhileration of watching the returns come in. But it is very different to spend a whole day making routine use of that same test after it has been published, and still more different to spend a day making routine application of a test for which one has no personal affection. Much of the routine testing in clinics is done by persons of comparatively little training, who have no vital interest in the results and who are not qualified to make constructive use of their observations. And if their work is made unreasonably difficult, it is conceivable that some of them will lighten the burden by taking unwarrantable liberties with the technique.

III. SUGGESTED CRITERIA FOR CLINICAL TESTS.

It is much easier to point out defects in the tests we are using than to devise methods for correcting them. Most of my recommendations are negative, but it can do no harm to call attention to the particular needs of the clinic.

1. The task or problem should possess sufficient intrinsic in-

---

[3] The inappropriateness of the Stanford Revision subject matter, as applied to adult subjects, has been discussed by Wells and Kelly, *Intelligence and Psychosis*, American Journal of Insanity, July, 1920.

terest so that the subject will usually regard successful performance as an end in itself. Spontaneous co-operation on the part of the majority of subjects should be considered essential. It is not to be expected that any test will meet with universal approval among clinical subjects, but there should be no tasks which are commonly performed only in response to urging.

2. Any test used for examination of adult subjects of whatever grade of mentality should possess sufficient dignity so that it can be offered to a mature, self-respecting man without insulting him. Questions should relate to politics, travel, the industrial world and other subjects of universal interest rather than to school life. All material that is essentially puerile should be rigidly excluded.

3. The requirements for each task should be simple and easily understood so that it may not be necessary to tax the attention of the subject by a long and involved explanation. Repetition should not be necessary and the time required for presentation should be negligible.

4. Questions which call for oral response, if used at all, should be satisfied by very simple answers.

5. The administration of the test should not be so exhausting as to make it impossible for an examiner in ordinary health to put in a full day's work of evenly good quality. Progressive deterioration in the quality of presentation should be avoided by the use of tests that can be presented with the minimal amount of actual labor. There should be very few tests, if any, which require the reading aloud of long paragraphs, and very few performance tests which call for close observation of the subject's movements. No tests should require hair-splitting distinctions as to what constitutes an erroneous movement.

6. Time limits are necessary in order to prevent an examination from being too long and tedious. But the time allowance in most tests should be liberal enough so that the subject will not be conscious of working under a time limit. Many normal persons are unable to do themselves full justice when working under high pressure, and it is unfair to class a person as deficient primarily on the ground that he becomes disconcerted when hurried.

7. Tests which call for reading knowledge on the part of the subject have a relatively small place in the clinic. Adult subjects who can read a little are usually very sensitive about their inability to read fluently, and the exposure of the deficiency in reading knowledge may be so disturbing emotionally as to invalidate the rest of the examination. Also, there are many clinical subjects who cannot read at all.

8. Tests which elicit obvious failures should be avoided as far

as possible. Such errors as are necessary to make a test discriminative should be interspersed among successful performances, so that the subject will not be discouraged by the sense of failure. Tests in which the subject will be wholly unaware of his errors are especially valuable. (Healy Pictorial Completion II furnishes proof that this is not impossible. A subject may make ten erroneous placings which net a minus score, and yet be perfectly satisfied with the result).

9. A scale of tests intended for clinical use should be of rather loose construction, so that thoroughness of the examination may not be dependent upon any particular test or group of tests. The system of scoring should permit giving the subject a rating on tests that are well suited to his command of language and his ability in general. It should not be necessary to give him a test that is hopelessly beyond him merely to prove that he cannot pass; nor, on the other hand, to give a test that he will take as an insult to his intelligence merely to prove that he can pass. The general plan of standardization for performance tests in which success or failure is obvious to the subject should be such that the examiner may permit or assist a subject to finish a task after the time limit is passed, so that the subject need not know of his failure to pass the test. And the rules for administration should be liberal enough to permit the examiner to select tests in accordance with the interests and ability of the subject.

Other criteria, of course, are essential. But it is not necessary to enumerate here the various standards to which educational as well as clinical tests are expected to conform.

The development of a scale which shall satisfy these requirements is a problem for collective rather than individual effort. However, individual contributions are not wholly at a discount, and at worst the problem is difficult enough to be intensely interesting.

IV. A TENTATIVE COMBINATION SCALE

This is offered somewhat prematurely, as a preliminary study, in the hope that it will encourage the splitting of composite educational tests into their component parts and the publication of independent norms for the separate units.

This plan for using a miscellaneous group of tests as a substitute for a standard scale is essentially the method suggested by Pintner and Paterson[4] for deriving a mental rating from a group of tests that were standardized more or less independently.

It seems to me that we can make a closer approach to the true mental age—if there be any such thing—by basing the estimate upon the median of a number of independent mental ratings than by accepting as final the rating derived from any single system of tests. I believe further that it is an advantage, other things being equal, to use tests drawn from different sources, developed by different authors, and standardized by results obtained from different sets of children. In this way the total number of normal subjects contributing to the norms is greatly increased, and certain constant errors, such as the personal bias of the experimenter, tend to be counteracted.

---

[4] *A Scale of Performance Tests*, chapter VI.

It is only as a tentative measure that I am willing to accept the term "mental age" as the end result of the series of tests. I consider the term absurd as applied to adults, and believe it to be frequently misleading as applied to children. It is far too readily understood, and is therefore used loosely by persons without psychological training. Worst of all there is a strong tendency to pass it around carelessly among workers in the clinic, sometimes within the hearing of a subject who is capable of comprehending its meaning. But although I strongly disapprove the use of this term and greatly prefer the percentile method, I find it necessary to express results of different tests in a term that will serve as a common denominator, and "mental age" is unfortunately the most convenient term for this purpose. My present aim, therefore, is to collect a large number of short-time tests which yield independent mental age ratings.

An abbreviated form of the Stanford scale—usually two or occasionally three items for each year—furnishes one unit of the series. Healy Pictorial Completion II is an exceptionally valuable test for English-speaking subjects of high grade. Of the fifteen tests of the Pintner-Paterson performance scale, I find the following especially useful: Mare and Foal Picture Board, Seguin Form Board, Five Figure Board, Two Figure Board, Casuist Form Board, Triangle Test, Diagonal Test, Ship Test and Substitution Test. Other tests which I am using already or expect soon to add to my collection are as follows: Kohs Block Design Test, Holley Vocabulary Test, Ferguson Form Boards, Cube Construction Test, Dearborn Form Board, Pressey Primary, Pressey Intermediate, Otis Intermediate, and selected tests from the Herring Revision of the Binet-Simon Scale.[5]

The essential feature of the programme is that each task presented to a subject can be selected with special reference to the interests and ability of this particular subject. Any test which does not elicit good co-operation is to be dropped, and the record, even if complete, is not to be counted in the series.

---

[5] I wish to express my gratitude to Dr. J. P. Herring, who has supplied me with records from which I can compute age norms for any tests which I wish to use.

The tests used in any examination should be as varied as the subject's ability and interests permit, and it is important to have a larger collection from which to draw, especially of verbal tests. If it were necessary to choose between language tests and performance tests I should give the preference to performance tests, because nearly all clinical subjects co-operate more readily in these tests than in tests requiring an oral response, and because there are many subjects who cannot be reached at all by language tests. But performance tests are limited in their scope, and many of them are highly susceptible to the influence of chance. They do not take the place of language tests, for subjects who have fair command of the language and the frame-work of the method will not be complete until verbal tests are available in much greater variety. The verbal-performance tests which are now being used extensively in schools offer a wealth of material for this purpose. The Terman Group Test, for example, contains several units which would be exceedingly useful in the clinic, given as individual tests. It will be necessary, of course, to compute norms for the separate units before they can be used in this scheme.[9]

It is my plan to use in approximately equal proportions tests which depend upon accuracy and tests which are scored by speed. While it is of course impossible to distinguish between these two variables in a strict sense, I think we are justified in disregarding one or the other in practice. In those performance tests in which the performance leaves no record of the errors, the difficulty of counting the subject's moves introduces a source of inaccuracy which it is particularly desirable to eliminate, since there is no way of checking up the examiner's count. In the Pintner-Paterson tests which I am using most freely, the error curve runs almost parallel to the time curve. It is true that the time record does not tell the whole story, but I cannot see that the count of moves or errors adds anything to it, and I think the examiner's time can be spent more profitably taking notes for qualitative study of the subject's mode of procedure. In giving the Kohs test recently to adult subjects of very superior ability, I observed that some toyed aimlessly with the blocks, thus making many useless moves; some worked by trial and error, moving so rapidly that I was unable to count the moves with even an approach to accuracy; and some worked very deliberately, with almost no false movements. All gave high scores, and the number of moves depended, apparently, more upon temperament than upon the ability measured by the test.

---

[9] I am indebted to Dr. Luella Cole Pressey for providing me with age norms for the four parts of the Pressey Primary test. This furnishes four excellent verbal-performance tests for subjects who understand English but who cannot read.

The mental ratings obtained from a given subject by different tests usually vary quite widely, sometimes ranging from sixteen years to zero. This may be due in part to the influence of chance upon some of the mechanical performance tests. But as the ratings which are affected by chance tend naturally to approach the extremes of the series, they do not invalidate the final result. In a series of fifteen tests it is not uncommon to find the middle four or five figures almost identical. If the ratings are fairly uniform I consider a series of ten tests sufficient, and I have occasionally assigned a mental age on the basis of nine tests. Ordinarily I prefer to make at least twelve tests.

The mental age as finally recorded is the median mental rating of all the tests—never less than nine—which are clearly discriminative for the subject, exclusive of any tests in which he does not give good co-operation and any tests which are too easy or too difficult to measure his ability. In the Manikin test, for instance, the highest possible score has a mental age value of eight years. If this record were counted in the series of a subject rating over eight years, the median would be lowered one point, and it would be absurd thus to penalize a perfect record. A test so difficult as to yield a zero score would have exactly the same effect upon the median, and it would be possible in testing a subject of low grade mentality, to give enough too-difficult tests to bring the median down to zero. A zero score does not necessarily prove the test to be non-discriminative, but I count the test too difficult to be discriminative when the subject works persistently up to the time limit without making any appreciable progress toward success. The examiner's errors in selecting the tests ought not to be counted against the subject's mentality, and it is clearly unfair to include these records.

By beginning with a test having a wide range of discriminative capacity and gauging the subject's ability by the results of the first test, it is usually possible to avoid giving non-discriminative tests. For a subject of low grade mentality I usually begin with the Seguin form board or the Mare and Foal picture board. For a subject of high grade the Kohs Block Design test is very satisfactory for the first presentation, because of its wide range of applicability and especially because of its unusual attractiveness as a toy.

SUMMARY

An indefinite series of short-time mental tests, verbal and performance, is recommended for use in the clinic as a substitute for a fixed scale of tests. Any short test for which an independent age norm is available may be used in the series, and the mental age is derived by taking the median rating of ten to fifteen tests which are clearly discriminative for the ability of the subject.

Many composite educational tests which are not suitable for clinical use in their entirety contain parts that would be highly valuable as single units of a series. The publication of norms for small homogeneous units, in addition to the norms regularly published, would be conducive to the wider usefulness of tests intended for educational surveys.

The following advantages of this method over the Stanford Revision have been observed:

1. The method is more elastic than a fixed scale, more adaptable to the interests of the subject. By selecting tests that are in line with his interests, it is possible to obtain better co-operation.

2. The method is adaptable also to the special qualifications of the examiner for administering tests, and the routine work of testing is thus made more attractive and less fatiguing.

3. The method is less mechanical. By allowing more room for common sense in dealing with a subject, it is conducive to more careful observation of the patient as an individual.

4. By placing less responsibility upon the examiner for the observance of petty rules, it reduces the chances that undue liberties will be taken with the technique of examination.

5. If the tests are well selected, it allows less play of personal equation in the scoring of results.

6. It is more economical of time. The period is spent upon tests that really test the ability of the subject in a given direction, and no considerable amount of time is wasted upon tests that are either too easy or too difficult to show definite findings. There is also an absolute saving in the time required for testing a subject of high grade mentality.

7. The validity of the results is not dependent upon the subject's co-operation in every test presented or in any particular test. If he can be induced to co-operate in any ten tests, it does not matter, for the validity of the test, how many questions he refuses to answer.

FRESHMAN TESTS IN THE SMALL COLLEGE[1]

By Elsie Murray,
Wells College

In this era of rapidly expanding college populations and changing standards in entrance requirements and curricula, the uses of a freshman intelligence test are manifold. In the smaller college admitting by certificate only, one may desire a criterion by which to gauge the quality of successive entering classes of entrants with and without Latin; or of students in the college in point with those in other institutions. The test findings are also utilizable in connection with student guidance, to detect special abilities or disabilities, and to throw light upon cases of academic failure. There is, besides, the possibility of discovering a test combination which may later prove of value in the selection of desirable (or the elimination of undesirable) candidates for admission.

With all of these objectives more or less in view, a freshman group intelligence test was given at Sweet Briar College, in the fall of three successive years: November 20, 1919, to 149 freshmen (class of 1923); October 26, 1920, to 116 freshmen (1924); and 29 seniors (1921); October 19, 1921, to 116 freshmen (1925) and 32 seniors (1922). The blank used was the Thurstone Psychological Examination for College Freshmen, in its 1919 form in 1919, and in the 1920 form in 1920. This test, which is arranged in 'cycle-omnibus' form, with a single time limit, requires only 30 minutes, and calls for no formal oral instructions. It was administered simultaneously to groups of about 40 each, by instructors coached by the writer.

Out of the various possibilities available in 1919, this particular blank was selected for trial, in part because of ease of administration, and economy of time and expense; largely, however, because the types of problem represented (*analogies, information, matching of abstract statements, judgment on the truth or falsity of arguments, sentence completion, completion of number series,* in the 1919 edition, the same with the substitution of an *extra word test,* and *arithmetical problems* for the two last mentioned, in the 1920 edition), had previously proved highly reliable in the writer's testing of college women (juniors and seniors).[2] As compared with Army Alpha, the Thurstone combination seemed better adapted to call out genuine and sustained thinking ability on the part of the student tested. In the long run, we must admit, the *sentence completion* test in the 1919 edition, and the *arithmetical problems* of the 1920, were found to constitute a serious obstacle to our use of this test. Not only does the substitution of the latter problems for the former invalidate the comparative value of test scores in successive years, but the wide variations in the possible solutions of the former (of quite different qualitative value), prevent any constancy in scoring from year to year. As regards the arithmetical problems, many students finding them to consume a disproportionate amount of time, simply omitted them entirely, thus running up relatively high scores for speed (problems correct) as compared with those of more conscientious individuals.

---

[1] *For aid and co-operation in the gathering of data for this paper thanks are due especially to President Emilie McVea, and* Dr. *Mary Harley, of* Sweet Briar College.

[2] E. Murray: *Psychological Tests as Diagnostic of Vocational Aptitudes in College Women.* Journal of Applied Psychology, March, 1920, pp. 33-34.

RAW RESULTS. TREATMENT OF SCORES. DISTRIBUTION.

Of the 448 students tested, only three covered the blank (168 problems in the 1919 edition, 182 in the 1920) within the time limit set. Among the seniors, no one scored higher than 152 correct out of 168, or 144 out of 182 among the freshmen. The test is therefore approximately long and hard enough for the group concerned, in the time allotted.

The pedagogical difficulty of handling students of widely varying mental calibre in a single section is a commonplace in education. In view of this, the range of scores within each freshman group is worth remark. The speed and alertness indicated by the upper scores (in *problems correct and problems covered*) is two to three times that represented by the lowest scores (see Table I). The range in proportion of errors and omissions to problems correct, as measured by *percentage accuracy*, is also striking, though less extended (56 to 96% for 1923).

The importance of calculating the percentage accuracy as well as the number of problems correct was first suggested by the fact that certain of the problems are couched in such a form that mere guesswork may yield a considerable number of successes; secondly, by scrutiny of the academic records of individuals in whom a relatively good score in speed or problems correct was observed to be accompanied by a relatively high proportion of errors. In many such cases, academic success or failure was found to bear a closer relation to the score in accuracy (number correct divided by number covered) than to the score in problems correct. E. g., in 1923, in more than half of the cases of striking dislocation between decile rank in grades and in tests, the discrepancy is considerably reduced by substituting rank in *accuracy* for rank in *problems correct* (see Table VIII). It seems not unlikely that while the latter may constitute the better measure of mental alertness and agility, in certain fields, the former (percentage accuracy) is a truer gauge of all-round ability, or at least of certain character traits-stability, perseverance, effort, ambition, and the related qualities, which enter so largely into the determination of academic success.

Accordingly, after considerable experimentation with scores and correlation coefficients, the method was adopted of computing an 'adjusted score,' by subtracting the sum of the errors and omissions from the number of successes. The coefficient for these 'adjusted scores' and freshman grades ranges from +.42 to +.44 for the different classes, and is always higher than that for the speed scores (e.g., for 1924, the correlation for academic grades and number of problems correct is only .375; for grades and adjusted scores, .42). Each class was then split into ten equal sections or deciles, on the basis of this score, those in the highest decile given a rank of 10, those in the lowest of 1, etc. Decile ranks were similarly determined for each student both in *problems correct*, and in *accuracy* (later, in *faculty estimates* and in *grades*) and a copy of the whole placed on file in the registrar's office for consultation by instructors and faculty advisers who desired further light on backward or promising students.

This adjusted score, it is obvious, raises somewhat the standing of the 'slow but sure' individual, (over the fairness of whose rating in a speeded test those unfamiliar with mental measurements feel such concern), as it also lowers somewhat the rank of the slap-dash, swift but careless; and there becomes thus available a set of three score rankings, which furnish jointly a more satisfactory mental picture of the individual than any one taken singly. The usual combinations of swift and inaccurate, slow and accurate, swift and accurate, and slow and inaccurate, obtain. While decile combinations of 6-1-3, 1-9-4, etc., for *speed*, *accuracy* and *adjusted* score, are not uncommon, and the correlation coefficient for speed and accuracy in 1923 was only .37½, there are also noteworthy cases of 10-10-10 and 1-1-1 (see Table II).

COMPARISON WITH FRESHMEN SCORES IN OTHER COLLEGES

The average score in problems correct for 1923 is 85.4. This falls a trifle below the corresponding averages quoted by Thurstone[3] for Liberal Arts and Engineering freshmen—86.6 and 88.8, respectively. The average for 1925, using the same test form,

---

[3] L. L. Thurstone: *A Cycle-Omnibus Intelligence Tests*, Journal of Educational Research, Nov. 1921, p. 265. The range of averages quoted for 34 Liberal Arts colleges runs from 61.6 to 107; *for 43 Engineering Schools* from 79 to 111.3.

TABLE I

DISTRIBUTION OF SCORES IN FRESHMAN TESTS, 443 SUBJECTS

	Freshmen			Seniors		
Class	1923	1924	1925	1921	1922	
No. Students	149	116	116	29	32	
No. Problems	168*	182**	168*	182**	168*	
Highest Score	168	169	158	182	168	No. Problems
75 Percentile	119	122	126.5			Covered
Median	108	107	113.5	119	138	
25 Percentile	94	93	98.5			
Lowest Score	63	60	62	83	105	
H	168	169	155	176	167	Attempted
Median	102	104	111	115	136	
L	58	61	61	83	97	
H	143	144	132	140	152	Correct
75	96	91.5	102	126	122	
M	83	80	90	98	108	
25	71.5	69.5	77	87	99	
L	49	42	53	53	84	
H	96	92	95	96.5	94.3	Percentage
75	84	84	87	88	88	Accuracy
M	80	76.5	81	82	84.8	
25	73	71	76	76	80	
L	56	43	60	55	62.5	
H	118	122	124	131	136	Adjusted Score
75	78	68	82.5	99	103.5	
M	60	56	67	73	85.5	
25	45	42	55	59	70	
L	12	—23	30	10	42	

* 1919 edition of test.

** 1920 edition of test.

rises to 90, a figure slightly higher than either of Thurstone's averages. There is, however, a weak spot in these comparisons,—the Sentence Completion test, which occurs in the 1919 edition, and the scoring of which it has proved impossible to standardize fairly.

MEASUREMENT OF IMPROVEMENT IN SUCCESSIVE ENTERING CLASSES

In the use of the tests to measure the mental calibre of successive freshman classes, the scoring difficulty above mentioned again arises. However, the scoring standard for sentence completion in 1921 was apparently more severe, operating if at all to lower the score for the class of 1925, since the median number of correct answers for this type of problem remains stationary for this class, while the medians for all other problems rise a point or two.

TABLE II

		BEST RECORDS			WORST RECORDS		
		No. Correct	Accuracy	Adjusted Score	No. Correct	Accuracy	Adjusted Score
1923	Score	120	96%	115	52	56.5%	12
Decile Rk.		10	10	10	1	1	1
1924		135	91	122	56	46	–10
		10	10	10	1	1	1
1921 (Senior)		136	96.5	131	53	55	10
		10	10	10	1	1	1
1925		131	95	124	62	67	30
		10	10	10	1	1	1
1922 (Senior)		152	90.5	136	105	62	46

A second hitch arose from the altered form of the test offered in 1920, the arithmetical problems of which proved a stumbling block to our students, many of whom possess a veritable 'number phobia.' The median score for this year drops in consequence 3 points instead of rising, although the general academic record of this class is above that of 1923. A certain degree of improvement in 1925 as compared with 1923 is, however, clearly discernible (see Table I and IV); median and lower limits are both higher, though no individual in 1925 attains as high a score as the best in 1923.

In Tables III and IV, the relation of senior and freshman groups is presented. Since the best senior scores barely exceed the best freshman, the claim of the intelligence test to measure native endowment rather than acquired ability or knowledge would seem to be substantiated.[4] Only the information and analogy scores of the seniors ascend disproportionately; the latter perhaps because of familiarity with similar tests gained in Psychology I. The bulk of each senior class ranks in the highest quartile of the freshman class subjected to the same edition of the test. This is, of course, as it should be, through the natural process of selection and elimination of the weaker operating over three college years. There remain, to be sure, in the senior classes a number of stragglers in the lower quartiles whose academic status is for the most part as unsatisfactory as their intelligence rating.

In order to check the operation of selection versus training more exactly, in January, 1923, when the freshmen tested in 1919 were seniors, the number of survivors in each of the original quartiles was tabulated (see Table III). The percentage of the survivors ranking above the median was found to be only 61 per cent, a figure somewhat below that for senior-freshman comparisons when two classes are tested simultaneously, the former in their senior year (84%). The 25, 50, 75, 100 percentiles for 1923 seniors were also found to average only about 5 points higher than those for the same group as freshmen, although the discrepancy between senior and freshman medians for 1922 and 1925 (tested simultaneously) was 18 points. Granted that the quality of entrants has not deteriorated from 1917-1921, individual improvement would seem to play a more important role than selection in raising senior scores.[5] Even so, the rise to be credited to senior superiority in training—in knowledge, aplomb, concentration, grasp, or what not—say 10 or 12 points on the average, forms only a fraction of the total range of difference in abilities in the original group (90 to 100 points). In other words, three years' college training would avail, presumably, to shift the rank of a given individual hardly a couple of deciles—a conclusion borne out by the records of retesting in the sophomore year, where the average gain approximates a decile.

---

[4] The fact that scores vary on the average inversely with the age of the subject, from 10 to 21, might also be cited as evidence that the test measures, not amount of training, but native ability.

[5] Although, before proceeding to any such conclusion, the unusual losses in the upper test and grade deciles of 1923—largely through transfer to other institutions—should be drawn into the calculation.

TABLE III

RATING OF SENIORS WITHIN THE FRESHMAN QUARTILES. DISTRIBUTION

	1921	1922	1921	1922	1923	
	For Problems Correct		For Adjusted Score		Correct.	Adj. Score
Quartile*	Frequency					
IV	17	21	18	20	13	11
III	7	9	7	7	10	10
II	2	2	1	4	8	9
I	3	....	3	1	3	4
Total	29	32	29	32	34	
Above median,	24 or 83 pr ct.	30 or 93 pr ct.	25 or 86 pr ct.	27 or 84 pr ct.	23 or 66 pr ct.	21 or 61 pr ct.

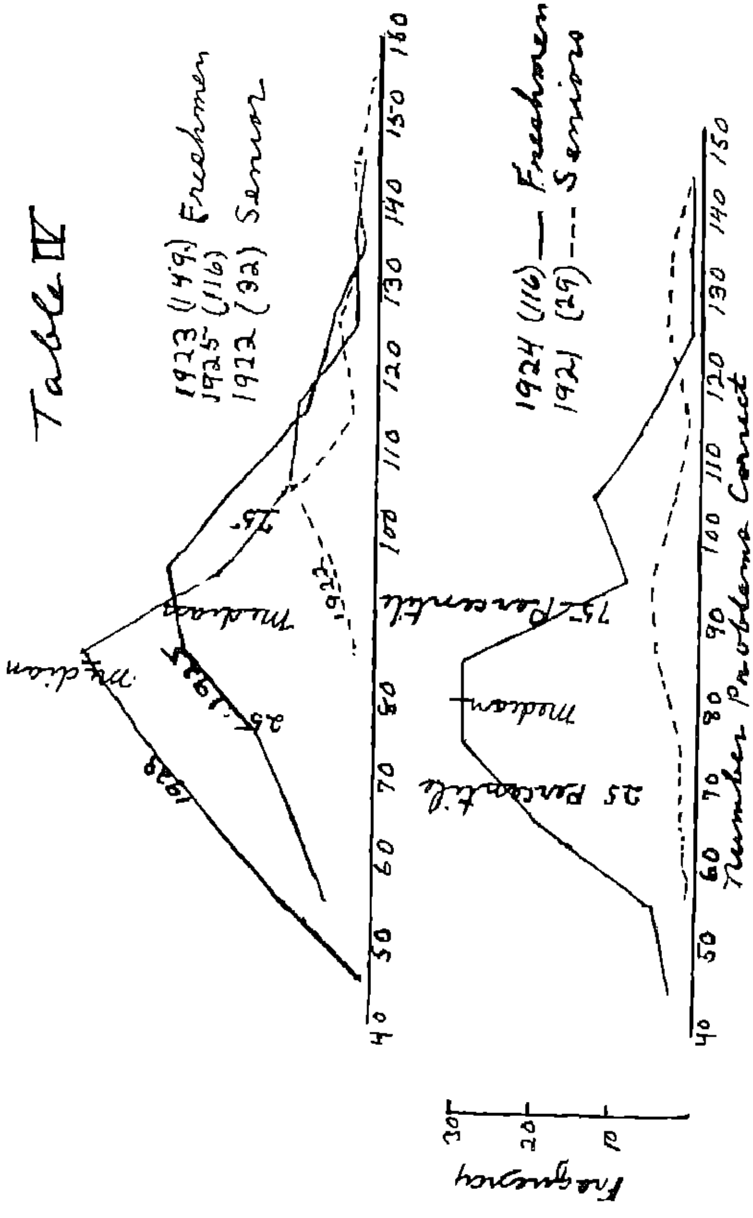
*Quartile rank in terms of 1924 freshman records for 1921, seniors
in terms of 1925 freshman records for 1922, seniors
in terms of 1923 freshman records for 1923, seniors

LATIN ENTRANCE CREDITS AND INTELLIGENCE RATING.

In order to estimate the possible detriment to the student body of altering the conventional Latin entrance requirement of the eastern colleges, both intelligence rating and academic standing for the freshman and sophomore years of students whose entrance certificates showed respectively 4, 3, 2, 1, or 0 Latin credits were tabulated. As shown in Table V, the median intelligence rating of those offering the standard number of units (4) is only slightly above the rating for those offering fewer. Further, the individuals of the two groups are similarly distributed over the ten test deciles, and the ratio of students above and below the median is practically the same (1 to 1) in each division. Adhering to the old 4-point Latin requirement would have cost the college 6 out of 15 of the most promising members of 1923; 4 out of 11 of 1924; and 2 out of 12 from 1925. If the intelligence test is a reliable indicator, the 4 unit Latin requirement is no longer the sieve for eliminating the weaker students which it seemed to be ten or fifteen years ago, when the brighter minds in the secondary schools were urged to complete the four years of Latin in the classical course, and failure to do so was a fair indication of either sloth or stupidity. This requirement now cuts out the sheep along with the goats.

From the point of view of academic success, at least in the sophomore and freshman years, the comparison of the reduced credit with the standard group *is less favorable. The median* grade for the former is somewhat lower, the number of students above the class average proportionately smaller. At the same time, the students of this group range through all the deciles of grade rank, and their elimination through a 4-point entrance re-

Table IV


1923 (149) Freshmen
1925 (116)
1922 (92) Senior
1924 (116) —— Freshmen
1921 (29) - - - Seniors
Median
25
1925
1922
75
1929
Median
25 Percentile
75 Percentile
40
50
60
70
80
90
100
110
120
130
140
150
160
Number Problems Correct
10
20
30
Frequency

quirement would have cost the class of 1923 5 out of its best 13 students; 1924, 1 out of 13; 1925, 3 out of 11.

Assignment of the reasons for the slightly lower academic achievement of the Sweet Briar student with reduced Latin entrance would be impossible without a fuller analysis of the conditions governing admission than is here practicable. There is, however, no real evidence that it indicates that the lack of Latin handicaps the student in either linguistic or scientific studies.[6] It may indeed signify that the steadfast pursuit of Latin studies is still in a degree a measure of the scholastic ambition and energy of the special group of students who apply to Sweet Briar for admission. This latter view would, in fact, seem to be supported by the falling away of 3-0 point Latin entrants in the sophomore year, as compared with 4 point—see Table V, in which the ratio of the former to the latter drops from roughly 1-1 to 1-2 for 1923, from 3-5 to 1-2 for 1924, in the sophomore years as compared with the freshman. This falling off is not, however, to be ascribed entirely to lack of scholarly ambition. The inability to obtain an A.B. degree on less than the full Latin entrance requirement operates either to discourage students from continuing in college, or to encourage their transfer to other institutions, e.g., state universities, where the lack of Latin is not a handicap. The effort

TABLE V

LATIN ENTRANCE CREDITS AND INTELLIGENCE RATING

FREQUENCY OF CREDIT GROUPS

Class	No.	4	3	2	1	0	Tot. 0-3	Credits Year
Fresh- 1923	149	79	18	32	5	15	70	
man 1924	116	71	14	24		7	45	
1925	116	73	17	22		4	43	
1923		46					27	Sophomore
1924		44					20	
1925		50					32	
CREDIT UNITS AND INTELLIGENCE RATING								
Median Test Deciles								
1923		6	5	6	4	6	5	
1924		6	5.5	4		6	5	
1925		6	5	6		4.5	6	
ACADEMIC RANK								
Median Grade Deciles								
1923		6	3	4	4	6	4	
1924		6	3	5		4	4	
1925		7	4	5		3	4	

---

[6] The English grades for the 1 unit Latin matriculates of 1925 actually average lower than those for the 2 unit group.

to make up the missing Latin credits in college in order to enter for the A.B. degree while carrying full freshman work also operates to lower their academic status. In any case, it is obvious that the effect upon the intelligence level of the student body of lowering the Latin requirement has not been at all as dolorous as the Latin enthusiasts were wont to predict.

TABLE VI.

CORRELATION OF TESTS AND GRADES

Year	Freshman	No.	Soph. and Freshman	No.	Jun., Soph. and Fresh.	No.	Jun. and Senior	Four Years
1923	.43	146	.43½	78	.517	38		
1924	.42	112	.48	65				
1925	.44	113						
1921	.34	21	.39				.48½	.48

PREDICTIVE VALUE OF THE TEST

Throughout the above discussion, the value of the Thurstone test as an indicator of ability has been more or less assumed. It remains to examine its claim more minutely, to determine the exact degree of reliability of the test in relation to success or failure at Sweet Briar College. With this object in view, the test scores have been correlated with instructors' estimates,[7] and with academic grades for from 1 to 3 years. The test ratings of students dropping out and remaining in each class have been compared. A number of members of 1923 and 1924 have been re-examined with an alternative form of the test. Cases of marked discrepancy between scores and academic ranks have been carefully analysed, and the health records of each class scrutinized for whatever light they may throw upon the problem.

---

[7] This method of testing out the scores was abandoned on account of the difficulty of securing faculty estimates of intelligence on any other basis than that of class marks in the department concerned. The correlation of tests and estimates for 1923 was .43½, only ½ point higher than that for tests and grades.

[8] Rogers, reporting 182 Goucher freshmen, finds a correlation of .37½ for the Thorndike tests and freshman grades (see Journal of Educational Psychology, April, 1920, p. 181; Mental Tests as a Means of Selecting and Classifying College Students). The correlations between grades and Army Alpha reported from Yale, Ohio, and Syracuse Universities range from .20 to .41. Higher coefficients are occasionally reported for special tests.

Test Scores and Grade Rank. After trying out and discarding various methods of weighting, the following procedure was adopted. The letter grades of each student for the year were converted into numbers (following the scheme A=11, A−=10, B+=9, etc.) and the total obtained. This total instead of the usual average was then made the basis of a grade ranking, and correlation coefficients with test rating (based upon adjusted score) obtained. These coefficients (R, calculated by the Spearman foot-rule, ×1.5) are shown in Table VI. The figure for the first year grades is about the same for all classes, +.43, and approximates that usually quoted for freshman groups of fairly homogeneous material.[8] When sophomore and junior grades are added, this figure rises to .517 for the 38 surviving members of 1923.

A clearer notion of the degree of correspondence (or divergence) between test and grade rankings may be obtained by dividing each class into four sections or quartiles, first on the basis of their test ratings, second, upon that of their year's grades. The number of students occupying a similar quartile in each set of measures may then be computed, also the number occupying the highest quartile in one and the lowest in the other, i. e., the number both of coincidences and of absolute disagreements in rank. E. g., as shown in Table VII, 22 individuals out of the 35 in the highest test quartile for 1923 rank also highest in freshma nscholarship; 14 out of 28 for 1924; 15 out of 28 for 1925. At the same time the range of academic achievement for any given degree of test performance is wide, and a few individuals from the lowest test quartile in each class attain the highest grade quartile, at least in their freshman year. The percentage of the group maintaining the same quartile rank in both sets of measures is 43 for 1923, 42 for 1924, 39 for 1925.[9] This means that we would be absolutely right in the case of about 4 out of 10 students in predicting their quartile rank in scholarship on the basis of their quartile position in test rating; absolutely wrong in the case of only 3 or 4 out of 100 (these latter discrepancies being reduced or eliminated when the ranking is based upon the grades of latter years.)

---

[9] The corresponding figure given by Rogers is only 35%. Whether this indicates the higher reliability of our scores, or the less selected and homogeneous character of our group, is debatable.

TABLE VII.

Quartile Coincidence In Test And Grade Rankings

A. 1923 FRESHMAN GRADES

		IV	III	II	I
	IV	22	7	5	2
TESTS	III	9	13	7	11
	II	2	8	12	9
	I	2	9	11	15
No. Subjects		35	37	35	37
Total					144
Coincidence					62(43%)
Disagreement					4(2.7%)

B. 1923 JUNIOR, SOPHOMORE AND FRESHMAN GRADES

		IV	III	II	I	Total
	IV	5	5	0	0	10
TESTS	III	2	1	6	4	13
	II	1	3	3	1	8
	I	1	1	1	4	7
Total						38
Coincidence						13(34%)*
Disagreement						1

C. 1924 FRESHMAN GRADES

		IV	III	II	I
	IV	14	8	5	2
TESTS	III	5	11	7	5
	II	8	4	8	7
	I	1	5	8	14
No. Subjects		28	28	28	28
Total					112
Coincidence					47(42%)
Disagreement					3(2.6%)

D. 1925 FRESHMAN GRADES

		IV	III	II	I
	IV	15	7	4	2
TESTS	III	5	9	9	6
	II	4	10	7	8
	I	4	4	6	13
No. Subjects		28	30	26	29
Total					113
Coincidence					44(39%)
Disagreement					6(5%)

---

*Raised to 44% when the test quartiles are equalised, on the basis of the 38 remaining members of the class.

E. 1921 FOUR YEARS GRADES

		IV	III	II	I
	IV	2	0	2	1
TESTS	III	2	3	0	0
	II	0	1	3	1
	I	1	1	0	4
		—	—	—	—
Total					21
Coincidence					12 (57%)
Disagreement					2 (9.5%)

For the test ranking of those delinquent in studies at the end of the freshman year, also of significance here, see Table XI. For a corresponding group Thurstone reports delinquents to the number of 0, 2, 3 and 9, for the test quartiles 4 to 1.

What is the real significance of these figures? Can a test which yields a correlation coefficient of less than .80, a percentage quartile coincidence with grades of less than 75, be of utility in a college office, either in the guidance or elimination of students? Should material or method be modified, or is the quest hopeless ?

In answering this question, the necessity must first of all be *emphasized of considering the grades for the entire college career* of the student before passing judgment upon the test as prognostic of academic achievement. Grades, always tricky, are specially open to attack in the freshman year. First of all, there is the *fallibility of the instructor*, possibly too easily impressed by a glib memory, poise of manner, a fluent tongue. Second, the *variance in ranking standards* in different departments, illustrated by the fact that the median freshman grade in the various subjects ranges from C minus to C plus for 1924, from C plus to B minus for 1925 (C=Fair, D=Conditioned, B=Good); while the middle point in one subject rose from C minus in 1920-21 to B minus in 1921-22. Third, owing to a certain *latitude of choice in subjects*, even in the freshman year, deceptively high records may be achieved by students doing very little real brain work; especially in the case of courses based upon material already familiar to the well-prepared student.

There is also to be considered the *routine character of many freshman courses*, involving on the whole more of a memory grind in grammar, formulae, and fact, than later, more advanced, and elective studies, which are *per se* better fitted to evoke real reasoning ability and initiative in the student. Lastly, from the point of view of the individual freshman, there is the *chaotic character of the first year* in college, involving as it does adjustments to a novel environment, collision of social and intellectual ambitions, as well as submission to a more or less arbitrarily imposed curriculum, in which the special bent of the individual may be ignored or thwarted. All these factors conspire to lower the value of freshman grades as prognostic of future academic success, and to prevent the individual student from finding her true level in the freshman year. An easy schedule or single-minded industry may run up a record unduly; the reverse factors lower it.

Even, however, when the upper class records are at hand, the indications are that the correlations will hardly rise higher than .60 or .70. Is this to be ascribed to the shortcomings of this particular test, or is it inevitable in all mental testing of college women, representing a comparatively narrow range of abilities and subject to certain other peculiar conditions? It is a matter of common knowledge that many factors apart from intelligence enter into the determination of academic success. Health, ambition, habits of industry, social distraction and diversion, worry and other causes may tip the scales one way or the other. Is it not likely, particularly in the case of a group of women, working less under the spur of strenuous vocational motives than do men, that the serviceability of any mental test is necessarily limited?

Analysis of Cases of Marked Discrepancy between *Test Scores* and *Academic Rank.* In order to apportion the blame between grades and tests, and determine if possible how far the test itself stands in need of alteration (in subject-matter, method of administering, or scoring, duration, etc.) a number of cases of marked dislocation between academic rating and test scores were selected for further study. Twenty-three of these, along with a sprinkling of seven others to act as a control—to the number of 30 in all for the classes of 1923 and 1924—were re-examined in March 1921 with alternative forms of the test blank. Secondly, a number of facts were collected and tabulated (see Tables VIII and IX) with regard to all freshman in whom the amount of dislocation is equivalent to at least ⅓ of the total range of rank for the group (representing nearly a quarter of each class.)

A considerable amount of the divergence between the two sets of ratings proceeds undoubtedly from inadequacies in grade ratings, the shortcomings of which, especially in relation to the freshman curriculum, have been already commented upon. As shown by the table, when sophomore grades are substituted for freshman, the discrepancy with test ratings is materially reduced; as a matter of fact, in three quarters of all such cases remaining in college. Cases of apparent overestimation by the test, rise on the average 2 deciles in grade rank; cases of underestimation drop 3 deciles.

Nevertheless, certain of the total number of cases in which the student's academic record exceeds the promise of her test performance, are probably *bona fide* instances of *underestimation by the tests.* Seven out of ten of such cases when retested rose several deciles in rank (although in the majority of control cases there was little change); and there is every reason to believe that the second testing yields the truer measure of their real ability. Various familiar strictures as to the fallibility of test data seem to be here substantiated—notably the objection that a low score may be a function not of native stupidity, but of lack of training, of poise, self-control or self-knowledge, manifesting itself as nervousness, flightiness, poor judgment as to the best balance to strike between speed and accuracy, exaggerated confidence or undue caution. Further, a scrutiny of the health records indicates that a certain number of the cases of discrepancy studied were laboring under some more or less temporary physical handicap, such as recovery from a recent operation, which invalidated the earlier testing.

In cases of apparent *overestimation* by the tests, the health factor enters also, though in a different fashion. At least half of the individuals in 1924 and 1925 whose grades fall seriously below expectation (15 out of the 30), and practically all whose sophomore standing shows no improvement, have poor health records; while, on the other hand, the majority of cases in the class of 1925 in which the grade rating is surprisingly high have excellent health records. There is undoubtedly a quick nervous type of individual, capable of running up a high score in a half hour test, but lacking the physical stamina (and with it the ambition), for the sustained effort essential to a brilliant academic career. In such cases, the high test score may sometimes be utilized to arouse dormant ambition, but the wisdom of so doing is often questionable; and such individuals are frequently outdistanced by classmates of mediocre ability but superior nervous and physical endurance.

While the uncertainty and variability both of health factors and mental attitude complicate thus the exact prognosis of academic success or failure from this or any similar test, certain devices are available for increasing the reliability of test findings. Pains should be taken to stress the importance of the test results to the individual (without creating undue tension); to defer the test period, in case of the physically unfit; and to repeat the test (in alternative form) a month or more later, at least for those occupying the lowest test deciles. Some more exact adjustment with regard to the inclusion of accuracy in the test score, or the weighting of errors, is also desirable. Something might also be gained by the supplementary use of a will test such as Downey's or a more complete physical examination record.

TABLE VIII.

DISCREPANCIES BETWEEN TEST AND GRADE RATINGS

Class of	1923		1924		1925	
Grade Ranking*	Higher	Lower	Higher	Lower	Higher	Lower
Number of Subjects	19	19	10	18	17	12
Median Age	18.3	17.8	18.4	18.3	18.3	18.1
Health Poor			5	10	9	6
Freshman Course Easy	13	8	0	10		
Left Before Sophomore Year	5	11	3	8	5	6
Soph. Grade Rating Nearer Test	11	7	6	5		
Accuracy Rating Nearer Grade	10	11	5	10	9	6
No. Retested	7	3	3	3		
Retested Nearer Grade Decile	4	2	3			
Total	38		28		29	

*As compared with Test Rating

TABLE IX.

RESULTS OF RETESTING

Class of	1923	1924	Total
No. Retested	16	14	30
Cases Discrepancy between Test and Grade	13	10	23
Control Cases	3	4	7
Second Test Ranking Nearer Grade Decile	6	5	11
Av. Decile Deviation of Freshman Grades from Test I Rating	4.9	4	
Av. Decile Deviation of Freshman Grades from Test II Rating	3.6	3.3	
Av. Decile Deviation of Sophomore grades from Test I Rating	3.8		
Av. Decile Deviation of Sophomore grades from Test II Rating	2.9		

It may be added that the factor of age, or rather, that of youth, seems also to complicate our results in a fashion not quite decipherable. There is a tendency to an inverse relation between age and test rank, a direct one between age and grades. The precocious mind tests high, but a certain maturity of experience and purpose as well as of mentality is apparently desirable in order that higher studies shall be assimilated in their full value, and studious habits set up. The weight of evidence is against lowering the age limit.

Test and Grade Rank of Withdrawals.

Another criterion of the academic value of test scores is sometimes sought in an analysis of the cases of withdrawal. For

Sweet Briar students, the distribution of the losses (shown in Table X) is fairly uniform over all the test deciles, except in the case of 1923, for which class, the mortality in the lowest decile is conspicuously greater than that in the highest. Again, the loss of students from the upper half of the test ranks is only slightly less than that from the lower—roughly as 5 to 6 for 1923, and almost 1 to 1 for the two later classes. Obviously test ranking cannot be used at Sweet Briar to forecast the probability of withdrawal as suggested by Thurstone. That this is not, however, to be interpreted as *per se* a reflection upon the reliability of test ratings is indicated by the second half of Table X: the proportion of total losses for the upper *grade* deciles practically parallels that for the upper test deciles—except in the case of 1925, where the record is as yet incomplete.

TABLE X
TEST AND GRADE RANK OF WITHDRAWALS

TEST RANK*

Year	1923 (154)***				1924 (116)			1925 (116)**	
	Fresh	Soph.	Jun.	Tot.	Fresh	Soph.	Tot.	Fresh.	Soph.
10th decile	2	5	1	8	4	4	8	4	1
1st decile	10	3	2	15	4	4	8	5	3
Above median	26	25	3	54	25	13	38	17	9
Below median	36	23	4	63	20	17	37	17	8
Total	62	48	7	117	45	30	75	34	17

GRADE RANK

	1923 (146)				1924 (112)			1925 (113)	
10th decile	3	3	1	7	2	4	6	3	0
1st decile	14	1	1	16	18	2	10	7	1
Above median	24	24	2	50	17	17	34	15	6
Below median	35	20	4	59	24	13	37	16	11
Total	59	44	6	109	41	30	71	31	17

*Adjusted score.
**Incomplete.
***Students registering at mid-years are included in this table.

TABLE XI
QUARTILE TEST RANKING OF DELINQUENT STUDENTS
(Those in lowest grade decile at end of freshman year)

Test Quartile	Test of 1923	1924	1925	Thurstone
4	1	1	1	0
3	3	2	2	2
2	3	5	2	3
1	8	3	7	9
Total	15	11	12	14
No. Re-registering	4	3	4	
Total number in class	149	116	116	150

The reason for this state of affairs is not far to seek. In a woman's college, manifold causes other than ability and ambition operate to reduce the ranks of re-registrants, among them ill-health, poor preparation, unwillingness to conform to college discipline, social ambitions, family affairs and finances, desire for specialised courses, or a whim for transfer to other institutions. In a smaller college such as Sweet Briar, many freshman enter with the deliberate intention of remaining only a year or two, and all the allurements of culture do not avail to move them from this intent.

Nevertheless, there remained in June, 1922, of the lowest test decile of the class of 1923, only one student; of the lowest of 1924 three only—all struggling under heavy handicaps and in the lower ranks of the class. In June, 1922, moreover, the honor students were all, with one exception (a student whose low information and arithmetic scores reduced her original record, but who ranked considerably higher in retesting with an alternative form of the test) drawn from the upper quartiles, if not always the upper decile of their respective classes. In January, 1923, there remains, out of the lowest decile of the class of 1923, no student; of 1924, two only; of 1925, three.

Lastly, if quartile coincidence in rank is reckoned between the academic ratings for the freshman and junior years of the 38 surviving members of 1923, there is found to be 37% coincidence, and 5% total discrepancy (i. e. two students, ranking in the lowest quartile in their freshman year, rise to the highest in their junior). The percentage coincidence for junior grades and tests is 39%, total discrepancies lacking.

Summing up, then, all the evidence, in face of the fact that at best barely a quarter of each entering class survives to graduation, it seems reasonable to conclude that the college would benefit materially by the lopping off at matriculation of the lowest test decile; and that the risk in so doing of losing any real college material is slight. Granted the precaution of a second testing, a dead line for admission might be drawn not merely at the 10 but even at the 25 percentile—a score of about 72 correct and 73% accuracy (in terms of the records of 1923 at Sweet Briar, for the 1919 edition), or better, at an adjusted score of 44. Where it is not deemed feasible to risk the reduction of the freshman class in this fashion, it would at least be desirable to encourage by every means available the students of the upper quartile, to discourage the return at the end of the freshman year of the students of the lowest Judging by the correlations and other data this course would in the end prove more satisfactory than merely seeking the withdrawal of the student of the lowest *grade* decile or quartile

The question of the suitability of the component elements of the two editions of the Thurstone test for our particular task, and the value of each relative to the different branches of the curriculum, is reserved for discussion in another place.

PREDICTION FORMULAE FOR TEAMS OF APTITUDE TESTS

By Clark L. Hull
University of Wisconsin

I

The rapid advance in the practical application of tests to the prediction of aptitudes, promises soon to bring into general use scientific prediction formulae. With the passing over of this important branch of test technique from the theoretical to the practical, comes a re-examination of the properties of the regression equation, now generally accepted as the standard prediction formula. It is the chief object of the present article to point out certain practical difficulties associated wtih the use of the regression equation for this purpose, and to propose a remedy.

Suppose the employment department of an industry has two jobs for each of which a team of tests has been devised and each team provided with an appropriate regression equation. Let it be further supposed that this employment department has been in the habit of rating its employees on a simple uniform scale such that the worst man in a hundred is scored zero, and the best scored 10, with the mean at 5; and that the diagnostic potency of the first team of tests (R) is .60 while that of the second is .40. Now if such an employment department should begin to use its tests and regression equations in the prediction of aptitudes of prospective employees and with exactly the same grade of men as previously employed, it would find that the highest prdicted score ever obtained with the first equation would be 8, and the highest with the second only 7, instead of the 10 to which it had been accustomed. If the empolyment department had been in the habit of discharging all employees known to be below score 2 in aptitude and had planned on rejecting all applicants who were indicated by the tests as probably below this level, it would find to its surprise that the first equation would predict none at all below 2, and the second none below 3! The equations would thus indicate no applicants as being either very bad or very good. If the employment department should desire to reject the lowest 25% of all applicants, it would have to select a radically different score from the score marking the lowest 25% of its actual employees. Not only this but it would have to select a distinctly different score to accomplish this with each of the two regression equations. Or, to consider the matter from the point of view of the order of merit, a predicted score of 6 by the first team of tests corresponds to a rank of *tenth* from the best in a hundred, whereas the *same* score with the second team corresponds to a rank of *third* from the best in a hundred.

Now the cause of all this confusion lies in the fact that regression equations have an inevitable and usually *radical* tendency to narrow the range of distribution of predicted scores as compared with that of the original criterion. This is at once evident from an inspection of the basic regression equation,[1]

$$x = r\frac{\sigma_x}{\sigma_y}y \quad \text{......} (1)$$

when it is recalled that r is always less than unity. The *operation* of this tendency in the case of multiple correlation may be shown by means of the following miniature problem, which will also serve to illustrate certain other principles to be brought forward later. The original data of the problem are given in Table I, together with the results of the necessary preliminary computations for the regression equation.

TABLE I

Subj.	Criterion (1)	Test A (2)	Test B (3)	Correlations	Partial S.D.'s
I	2	8	8	$r_{12}=+.625$	
II	4	5	5	$r_{13}=+.5625$	$S.D._{1.23}=2.78$
III	6	7	2	$r_{23}=+.375$	$S.D._{2.13}=1.56$
IV	13	9	14		$S.D._{3.12}=3.31$
V	10	11	6	$r_{12.3}=+.542$	
	—	—	—		
Mean	7	8	7	$r_{13.2}=+.452$	
S. D.	4	2	4		

Dropping the smaller decimals, the primary regression equation on this criterion is,

$$x_1 = .968x_2 + .381x_3 \quad \text{......} (2)$$

where the x's are the deviations of the individual subjects' scores from the means of the separate arrays. Substituting in this to obtain an equation in which actual test scores may be used for purposes of prediction,

$$X_1 - 7 = .968(x_2 - 8) + .381(X_3 - 7) \quad \text{......} (3)$$

Clearing and collecting, we obtain,

$$X_1 = .968X_2 + .381X_3 - 3.41 \quad \text{......} (4)$$

where the X's represent actual scores. This is the form of the regression equation ordinarily used for purposes of prediction.

---

[1] Yule, G. U.: *Introduction to the Theory of Statistics*, p. 171.

Substituting in equation (4) the test scores of each subject shown in Table I, we obtain by successive application, the following "predictions":

Subject	Predicted score	
I	7.38	
II	3.33	
III	4.13	
IV	10.64	
V	9.52	$R=+.719$
Mean	7.00	
S.D.	2.877	

These predicted scores correlate with the criterion to the extent of +.719. The mean of the predicted scores is seen to be, as always, the same as that of the criterion, 7. It will be noted, however, that the S. D., as would be expected from the nature of the regression equation, is much less than that of the criterion, being only 2.877 as against 4. This is what caused the confusion noted above.

II

As indicated by equation (1) the amount of shrinkage in the dispersion of the predicted scores is dependent upon the size of r. In cases of multiple correlation such as the present where we have a predicted criterion derived from a set of weighted tests, the shrinkage in dispersion is dependent upon the size of th correlation between the predicted and the actual criterion scores. Yule has designated this correlation as R.[2] Thus in the above example, if the S. D. of the criterion (which is 4) be multiplied by R (.719), the result will be 2.877, which is identical with the S. D. of the predicted scores as obtained by actual computation. Knowing this relation then, we may write the formula,

$$x_1 = cb_{12.3}x_2 + cb_{13.2}x_3 \quad \ldots\ldots (5)$$

where

$$c = \frac{1}{R}$$

and all of the other symbols are the same as employed by Yule. This formula, while leaving the correlation between the tests and the criterion undisturbed, yields a prediction series with exactly the same mean and degree of dispersion as the criterion itself, thus eliminating the objectionable features of the regression equation noted above. It is accordingly proposed for use in vocational and employment prognosis.

---

[2] Yule, op. cit. p. 248.

In determining R, it may be recalled that it is not necessary actually to make the predictions and then compute the correlation between prediction and criterion as above, in order to find R for use in formula (5). R may be found very readily by formulae of the following type:[8]

$$R_{1.23}=\sqrt{1-(1-r^2_{12})(1-r^2_{13.2})} \quad \ldots\ldots (6)$$

Substituting appropriately in equation (6) from the values given in Table I,

$$R_{1.23}=\sqrt{1-(1-.626^2)(1-.452^2)}$$
$$= .719$$

This agrees exactly with the r computed from the predicted and the original criterion scores in the ordinary manner.

The operation of the proposed prediction formula (5) may be illustrated by means of the example already considered. Substituting the value of R in the formula for c,

$$c=\frac{1}{.719}$$
$$=1.391$$

Substituting the value of both c and b in formula (5),

$$x_1 = 1.391 \times .968\ x_2 + 1.391 \times .381\ x_3$$
$$x_1 = 1.345\ x_2 + .530\ x_3 \quad \ldots\ldots (7)$$

Equation (7) is analogous to equation (2) above. From here on, *the procedure is exactly like that with an ordinary regression equation. Substituting the values of the means as in equation (3) and solving, we have,*

$$x_1 = 1.345\ x_2 + .530\ x_3 - 7.468 \quad \ldots\ldots (8)$$

Substituting the test scores of the various subjects in (8) as was done in (4), the following predictions are obtained:

Subject	Predicted Score	
I	7.529	
II	1.844	
III	3.007	
IV	12.051	R= .719
V	10.506	
Mean	6.987	(7.
S. D.	3.995	(4)

It will be observed that in this prediction, not only the mean is the same as that of the original criterion, but that the standard deviation is also the same. The shrinkage in the dispersion of the predicted scores caused by the regression equation has been corrected by the prediction formula (5), and the standard deviation restored to its original size of 4. The correlation is entirely undisturbed, standing at .719 exactly as before.

---

[8] Yule, op. cit. p. 218.

It is to be noted that the prognosis formula employed above is not itself a regression equation. The true multiple regression equation serves two distinct functions. One function is to give the weight that each test score should be multiplied by, that all combined shall yield the maximum correlation with the criterion. Innumerable such optimal weightings are possible, since all that is necessary for this is that a certain optimal proportion shall be maintained among the weights given to the various tests. The second function of the regression equation is to indicate the particular set of optimal weights that, when the test scores are multiplied by them and combined, they will yield the most probable *absolute* criterion score obtainable from the data. The proposed formula performs the first function but not the second.

To give the most probable absolute criterion score, the regression equation must always reduce more or less extensively the dispersion of the predicted scores below that of the criterion aimed at. This introduces an element of uncertainty into their use for employment purposes, very much like what would happen if predictions were to be made on a new and unfamiliar scale. Thus with a basic 10-point scale, the maximum possible predicted score may vary from 6 or so, up to above 9. For this reason a predicted score of 6.5, say, has little meaning when applied to an individual subject. It may mean that he is probably only slightly above the average applicant in aptitude, or it may mean that he is probably the one best applicant in a hundred. The predicted score, standing alone, gives us no light whatever as to which of these two extremes is more nearly approached. Unfortunately this is exactly what the employer wishes to know from a team of tests. By restoring all predicted dispersions to the size of the standard criterion aimed at, the proposed formula eliminates this ambiguity.

It is also easy to show that, in spite of the fact that the regression equation yields the most probable absolute predicted score, its ordinary use in employment may lead to unfortunate results. Let us revert to the employment situation instanced above (Sec. I). Suppose the labor market to be such that it is necessary to hire practically all applicants, the two teams of tests being used merely to decide in which of the two jobs a given applicant probably has the greater aptitude. Let it be further supposed that aptitude in the two jobs is positively correlated. One result would be that men of fairly equal but low ability on both jobs would tend to score higher on the second job, since the low R of this team of tests tends to raise all low scores towards the mean, thus tending artificially to shunt the poor workers into the second job. On the other hand, subjects having relatively equal but high aptitude for the two jobs will tend to score *lower* on the second team of tests because the low R in this case has lowered towards the mean predictions from all high scores, thus tending artificially to shunt the good workmen *out* of the second job. As a result of this there will be a tendency for the poor workmen to be segregated on the second job and for the better workmen to be segregated on the first job. But with the proposed prognosis formula, since the range of prediction would be exactly the same for both teams of tests, the unfortunate tendency noted above could not take place. It would still be true, of course, that the predictions made from the team of tests with the weaker R would be less accurate than those made from that with the stronger R.

III

It will not infrequently happen that the original unit of measurement of a criterion will be such as would be undesirable for use in prediction. For example, the original criterion of the efficiency of lathe workers might be a set of careful micrometer readings showing the deviations from specifications of certain products turned out by the subjects. Obviously such a criterian, while probably excellent from the point of view of reliability, would hardly be sufficiently intelligible for general use in the prediction of aptitudes[4].

A special need to make predictions in terms other than those of the original criterion measurement will arise where it is desired to predict the aptitude of a person on a number of different vocations for purposes of comparison, with a view of determining in which he is most likely to succeed. Such comparative predictions, it would seem, must become very general if scientific vocational guidance is to become a reality. To be readily comparable, such predictions should all be made in terms of the units of some simple uniform scale, e. g. one of ten points. This scale will ordinarily not (except by chance) be the same as that of the original criterion measurement. For such cases the following method has been derived by means of which predictions may be made in terms of scales having any desired means and any desired standard deviations, quite irrespective of what the corresponding values of the original criterion scales may have been.

---

[4] A rather common form of this difficulty requiring special treatment is where, as so often happens, the only available criterion is a set of ranks of the subjects in the activity studied. Ranks are not only unsatisfactory units for purposes of prediction, but significant regression equations cannot be based upon them. For such cases, a method of converting ranks into linear units convenient for purposes of prediction and upon which regression equations may be based, is described in this Journal Vol. VI. pp. 385 ff.

The procedure is first to work out the primary regression equation on the basis of the original criterion measurement exactly as in (2) above. Then if we let σ_1 be the S. D. of the original criterion scores, and σ_1' be the S. D. of the scores to be predicted on the desired scale by the prediction formula, we may write,

$$c = \frac{\sigma'_1}{R\sigma_1} \quad \text{.......................................} (9)$$

This value of c is substituted in equation (5). Lastly the mean of the desired prediction series is used in place of that of the criterion series when the x-values are converted into X-values as in equation (3) above.

The concrete working of the method may be illustrated by transforming equation (5) into one which will predict in terms of a 10-point scale with the mean at 5 and with a S. D. of 2. Substituting the appropriate values in (9),

$$c = \frac{2}{.719 \times 4}$$
$$= .698$$

Substituting the value of c in equation (5) and solving we have,

$$x_1 = .6729\ x_2 + .2649 x_3 \quad \text{.......................................} (10)$$

Substituting in (10) the values of the means as in (3) above except that here 5 (the desired mean of the new predicted series) replaces the 7 of the original criterion series, we have:

$$x_1 - 5 = .6729\ (x_2 - 8) + .2649\ (x_3 - 7)$$

Clearing and combining,

$$x_1 = .6729 x_2 + .2649 x_3 - 2.2375 \quad \text{.......................................} (11)$$

This is the formula that we have been seeking.

Lastly by substituting the test scores of Table I in formula (11) we obtain the following predictions, this time in terms of the desired 10-point scale:

Subject	"Predicted" Score	
I	5.265	
II	2.453	
III	3.003	
IV	7.528	r = .719
V	6.753	
Mean	5.000	
S. D.	2.000	

It will be observed that the prediction yielded the required mean and standard deviation, and that the correlation is still undisturbed.

The same results as secured above may be obtained quite simply by converting the undesirable criterion scores into the desired prediction scale before the zero order correlations are computed. This may be performed with ease and precision by a method described in this Journal, Vol. VI (p. 298 ff.)

In case it should be desired to predict in terms of a scale different from that of the original criterion but by means of a true regression equation instead of the prediction formula recommended above, the procedure is exactly the same as that just described, except that in this case,

$$c = \frac{\sigma_1^{1}}{\sigma_1}$$

LABOR TURNOVER AND METAL ALERTNESS TEST SCORES

BY A. J. SNOW
NORTHWESTERN UNIVERSITY

That labor turnover is an ever present problem and a costly one to the industrial institutions that harbor it and to the community generally is admitted by students. Some industrial managers fully realize the waste involved in the maintenance of an unnecessarily large labor turnover; if there is any one who doubts it, it is that group of employers who do not possess the capacity to move along mentally in the realm of industrial relations with the constant industrial expansion. It is well known that this larger group, composed mainly of inefficient and dogmatic employers, is the greatest single obstacle in eliminating this form of social waste. That this group of men is always willing to assert their *superior* and infallible knowledge in handling employees, is an ever present experience of the student in this field, and that they have almost completey failed in dealing with this problem as well as with the other problems of industrial relations, we have proof of in the increasing industrial unrest.

Labor turnover is not only an industrial problem, but directly a social and community problem, this can be seen when we realize that waste in any form adds to the actual cost of production, which is paid exentually by the community of consumers. To point out only five of the many items which are chargeable to this form of preventable waste, from the point of view of management, we have the following: first, expense incident to employing; second, instruction to the new employee by the foreman; third, spoiled work because of new surroundings, new tools, etc.; the same is true of the fourth factor,—breakage of tools and machinery; fifth, decreased production due to the temporary low production of new help. From the point of view of the ill-placed employee, we can enumerate the following three factors: first, loss in wages incident to the change of jobs; second, decrease in earnings due to the period of adaptability to the new job; third, decrease in earnings due to misplacement.

It has been known to industrial psychologists, some of the more enlightened employers and a handful of economists that wages and working hours are not the only industrial problems; in fact, wages and hours, as an incentive, are subordinate, or, at most, on a level with the control and distribution of power within an industrial organization. All that one has to do to verify this hypothesis is to estimate the number, magnitude and importance of disputes over the distribution of power as compared to that of wages and hours in any industrial court of arbitration of long standing, such as the industrial court in the clothing industry. The same may be said of our present problem—labor turnover. The casual observer is certain that the most responsible factor for an abnormal labor turnover is the worker's dissatisfaction with his wage. The student of the subject knows, however, that dissatisfaction with wages will not alone induce the worker to change his job, much less to give it up with the ever-looming possibility of unemployment. Most human beings must feel that the change will be beneficial in more ways than one. As a rule there are a number of factors contributing to the worker's decision to leave, and some of these may very well be of the following nature: abuse of power—perhaps the greatest source of dissatisfaction—wages, hours, unsanitary conditions, arbitrary advancement policy, lack of instruction, faulty machinery or tools, heterogeneous "consciousness of kind" among team workers, personal incompatibility with immediate superiors, lack of standardized methods of process, lack of instruction, lack of specialized training on the part of the worker, economic insecurity, physical incapacity, mental incapacity and a mentality too great for its task.

Statistics are unfortunately lacking as regards the comparative importance of most of these various factors. However, some interesting data has recently been gathered in connection with the two last factors. While admitting the danger of any study which aims to dissociate any one factor from its immediate surroundings, nevertheless, such a study may be justified on the grounds of its illuminating results and their actual use, when other factors composing the complete situation are given due weight. Consequently, the statistical study as to the relation of mental alertness and labor turnover that is to follow must not be given undue prominence in the mind of the student of the general and wider problem of labor turnover.

The Scott Company has found the following tendency in their practical application of psychological methods to industrial management—that, when all of the men in any organization, excluding major and minor executives, are given the Scott Company Mental Alertness Test[1], the score so obtained tends on the average to change with the worker's occupational group[2]. This tendency would suggest that a worker's place in an occupation which demands higher capacity for alertness than he is capable of would become either dissatisfied with his abnormally slow progress or would be discharged for incompetency. On the other hand, a worker of higher alertness capacity than is demanded by his present occupation, if not rapidly promoted into another field of endeavor where a higher level of alertness is necessary, will become dissatisfied and leave for a more promising job. The worker falling in this latter class is, of course, very valuable to an industrial enterprise and the knowledge of his capacity in mental alertness would tend to save him for that industry which gives due weight to such information in their promotion and training policy.

---

[1] Most psychologists recognize, and rightly so, a decisive difference between mental alertness and intelligence, mental alertness being only one factor in intelligence. See, e. g., Gault, *Social Psychology* (Holt, 1923), p. 77.

The following two experiments made by the Scott Company yield illuminating data. The first one was a study of the relation of the number of years each man was retarded at the time of leaving school—as an index of mental alertness—as related to the stability of the working force in several production departments of a manufacturing company. In this study there has been found a striking difference between departments, e. g., in the tool department where the work is of a high grade and varied, "the greatest proportion of dissatisfaction occurs among the workers who were most retarded in school. The stability increases as the amount of retardation lessens. . . ." While in the inspection department of the same company, where the work is repetitive, monotonous and 'fool-proof' "the amount of dissatisfaction is low for those men who were very retarded in school. The percentage of dissatisfaction increases markedly until, for those whose progress in school was normal. . . ."[3]

The second study dealt with relationship of mental alertness test scores and labor turnover. The psychological test used was the Scott Company Mental Alertness Tests.[4] This study has been made upon the employees of three different companies. For the sake of convenience we shall designate the three companies as X, Y, and Z. The test was given to the usual routine office clerks of company X and Y, the only difference being that women clerks were the subjects in company X, while men clerks in company Y. In the case of company Z office boys only were examined. In each case there has been found to exist a definite relationship between the mental alertness of employees, as determined by test scores, and labor turnover. However, this relationship is not always the same for the three companies.

---

[2] All of the freshmen at Northwestern University had to take the Scott Company Mental Alertness Test. It has been found that after the lapse of one semester that all of the freshmen that scored in the upper one-third in mental alertness tests 00.3 per cent earned a grade, which if averaged throughout their college career, would enable them to graduate. Of the freshmen who scored in the lowest one-third in the mental alertness tests, only 20 per cent secured the necessary grade for graduation. Freshman whose high school scholarship would place them in the fourth quarter, and who were admitted to the University on securing the medium or above in the mental alertness tests, only 16.0 per cent received the grade at the end of the first semester, which if kept up, would enable them to graduate. (The data for this note was secured from Mr. Paul Lester Palmer, Northwestern University.)

[3] Scott and Hayes *Science and Common Sense in Working with Men* (Ronald Press, 1921). p. 75. For a full account see also pp. 72-28, and the Scott Company Laboratory Bulletin L, No. 3, August 30, 1920. This bulletin is not published but mimeographed; the same is true of the bulletins referred to below.

[4] Printed, not published by the Scott Company, Philadelphia, Pa.

For company X, it has been found that about 40 per cent of the women clerks hired who have an alertness score between 15 and 30 leave within six months; this percentage decreasing rapidly with the increase in the test scores, until the least labor turnover is found among clerks who scored between 35 and 50. For those whose scores were above 50, the percentage leaving increases rather rapidly until we reach those scoring highest, where the instability is the greatest, greater than for the clerks scoring low in the mental alertness.[5] From the above relationship of mental alertness and turnover we may infer that company X does not give due weight to the necessity of rapid promotion for its most alert clerks.

For company Y, as in the above case of company X, the labor turnover is high for employees making low scores; instability rises with an upward trend of the alertness scores until the highest scores are reached, where there is found a pronounced increase in instability, a condition unlike that of company X. That difference between the two companies appeared, at the time, to be due to a peculiar wage situation causing instability among the middle-high employees.[6] That points again to our original contention regarding the cause of high labor turnover, that complete generalization is impossible when the phenomenon is studied in the light of only one isolated factor. What can be said, however, is that it is not true that the "make shift" or the "general run" of employees is responsible for the high labor turnover, but among other factors the incompatibility of the requirements of the job with the capacity for mental alertness as manifested by the worker, is an important contributing cause—and perhaps, *at times*, the only cause for labor turnover. That becomes of further significance with the multiplication of our experiments, such as the following one, made in company Z.

In company Z in February, 1920, one hundred and eleven boys under eighteen years of age were given the mental alertness tests. The results were filed away and no use was made of them until November, 1921, when the scores were brought out of the files and compared with the progress made by the boys during the intervening twenty-one months. It has been found that the average score for the one hundred and eleven boys was 38.7, while the score for the sixty-three boys, still in the company's employ, was 41.6, and the average score of those no longer with the company was 35.1. That the boys making a high score are most likely to remain with the company becomes evident from the above figures; while the boys leaving the company's employ did so for three reasons: first, for better positions—their average score was 44.5 (notice that the score is higher than that of the boys remaining in the employ of the company); second, to return to school; third, discharged by the company—their average score was 26.1.

---

[5] Scott Company Laboratory Bulletin L. No. 2, March 25, 1920.

[6] Ibid.

If we should compare the salary of the boys on November 1921. with the alertness scores we find the following relationship:

Number of boys.	Average alertness score.	Average Salary.
10	57.3	$16.00
10	23.1	$13.40

It has also been found that the average score, for the twenty-nine boys who were promoted to junior clerkship was 46.2, while the score 35.1 represents the average score of the thirty-four boys who have received no such promotion.

It is not only interesting but it is of importance industrially to notice that the boys who scored relatively high in the test either continued in the employ of company Z or left for better positions. Almost all promotion and, therefore, higher rate of pay were won by the boys taken from the same class, even though the alertness scores were unknown to anyone save the industrial relations department.

It seems to us that, while the results from the above experiment are too scanty to permit any positive generalizations, nevertheless they are suggestive of what might be done. Consequently, we may venture to make the following suggestions: first, that further study in this field might determine the relative alertness scores which are most suitable for some of the industrial occupations—where alertness has been found to be one important factor; second, that knowledge so gained will be of practical importance to the employment manager; third, that with this knowledge in the hands of the employment manager, he will be able not only to reduce labor turnover, where previously the instability was mainly due to the incompatability of mental alertness with the responsiblity of the occupation, but, also, he will be able to detect individuals of high alertness capacity whose services to the industry will be in their ability to learn rapidly and in that way gain quick promotion, ending finally with a job, the demands of which will be proportional to the worker's ability. The third suggestion expresses an ideal situation which we may put in the sky, as Plato did with his ideal state.

NOTES AND NEWS

HEALTH FELLOWSHIPS AND SCHOLARSHIPS

The American Child Health Association, believing that health is a matter of sound education, and that no greater creed can be spread through the world than the gospel of good health, has instituted an unique series of health scholarships.

The sum of $10,000 has been allotted to these scholarships and fellowships, which are to be given to teachers in service who have been especially successful in teaching health. The money will be distributed in five $1,000 scholarships and fellowships, to college graduates now in service as principals or elementary school supervisors; five $500 scholarships to normal school graduates or college graduates now engaged in classroom teaching; and fifteen $200 scholarships to holders of teacher's certificates now engaged in classroom teaching.

Those sums are not hard and fast ones, the Committee announces, but approximate, depending upon circumstances. It also announces that although the conditions require the teacher to be in service at present, an exception to this rule may be made in favor of any teacher who has already done exceptional work in health education, even if she is not engaged in teaching at present.

The purpose of these scholarships and fellowships is to improve professional training, to enable teachers to do more effective work in health education.

Fellowship and scholarships shall be awarded to *teachers in service* and to teachers not now in active service but who have had exceptionally successful experience in health education, who present evidence of some of the following qualifications:

a. Success in teaching.

b. Pleasing and favorable personality.

c. Good personal health.

d. Demonstration of unusual interest in health education.

e. Possession of initiative, originality, leadership and organizing ability.

f. In addition to the above, a good scientific background is considered of special value, in particular, instruction in biology, bacteriology, chemistry, nutrition, personal hygene, public health, child hygiene, school hygiene, child psychology and principles of teaching.

The qualifications of the Institutions in which fellowships and scholarships are to be placed should include the following:

1. Recognition of standards for the promotion of the personal health of their own students.

2. A healthful environment in the institution for the setting up of desirable standards from the standpoint of sanitation, etc.

3. A supplementary course of coordinated subjects of instruction such as biology, nutrition, etc., in addition to the more specific training in health education.

4. Opportunities for observation, preparation of programs, practical teaching in health education.

5. The providing of training for teachers in other fields than the field of health education.

6. A spirit of co-operation among the various departments connected with health education, as evidenced by their success in co-ordinating the various elements contributing to health education.

Fellowships and scholarships providing for a year's study shall go into effect with the beginning of the school year or in the fall of 1923.

Summer School, Chautauqua and Travelling Scholarships will go into effect during the summer of 1923.

The fundamental subject matter in any student's course should be derived from the following fields: Personal Hygine, Nutrition, Community Hygiene, Social Hygiene, Mental Hygiene, Health and care of infants and young children and adolescense, First aid and safety, Hygiene of the worker, Home nursing and care of the sick, School Hygiene, Psysical Education, Principles of Health Education and Practice Teaching. Practice Teaching to include practice in all types of contact with children incident to health work in the school.

For information address American Child Health Association, 370 7th Avenue, New York City.

BETTER EYE SIGHT AS A NATIONAL ASSET

The Eye Sight Conservation Council of America represents a distinct and organized phase of the general tendency towards social, economic and physical betterment. It aims to bring about improved vision among the masses of the nation, experience in education and industry having shown that a long train of ills, involving grave consequences, follow in the wake of defective eye sight.

One of the conspicuous services performed by the Council was its participation in the nationwide assay of waste conducted by Herbert Hoover as chairman of the Committee on Elimination of Waste in Industry of the Federated American Engineering Societies. This committee, whose field activities were directed by Mr. Wallace, found that heavy annual economic and physical losses were occurring in American industry through poor vision of the workers. All this, for the most part, was characterized as avoidable waste, which the Eye Sight Conservation Council has undertaken to wipe out by a thorough and systematic plan of education.

Of the 24,000,000 school children in the United States, approximately 10,000,000 can justly claim that their country in not giving them a square deal, according to Prof. Wood of Columbia. "For," he says, "they are laboring under a grave handicap—eye defects—which could be mitigated to a great extent, if not corrected entirely, by the intelligent supervision of the state and of the teachers of the nation, through proper eye tests of all pupils." Retardation is another evil caused in part, at least, by imperfect sight.

Experiments to determine the true economic value of perfect vision will be made. It is proposed to measure the improvement in health, increase in quality and quantity of production, advancement of individual performance, and decrease in losses due to waste and accident. Factory, home and school lighting, now a subject of scientific research here and abroad, will be studied.

The Council, through its Field Secretary, Charles F. Southard, is prosecuting its work in the schools of many cities. This work embraces lectures to parents, teachers and pupils with motion picture illustrations.

Co-operation with educational authorities is the prime purpose of the Council. An illustration of this co-operation is found in New York City, where the Council participated in the observance of Health Day in the schools, when the eyes of more than 1,000,000 pupils were tested.

Wherever the Council works, it strives to establish the principles of continued supervision of the children, mothers co-operating with the teachers in identifying defects, often hidden, and in applying corrective measures.

BOOK REVIEWS

J. W. W. WALLIN. *Causative Factors in Mental Inferiority and the Prevention of Degeneracy.* Miami University, 1922, 48 pp.

The writer gives a brief account of the factors causing mental inferiority. The article is a summary of his study of the personal and family histories of 872 clinic cases. The individuals studied are public school cases and not institutional or hospital cases, consequently the conclusions cannot be affected by hereditary cases as is so truly applied to institutional cases. After a thorough discussion of personal and family history data and with many carefully worked out tables, the author has drawn the conclusions that inebriety occurs more frequently than any other factor. He also makes a statement that the same thing applies to immorality. He summarizes other cases of mental deficiency, such as head injury, cerebro-spinal meningitis, infantile paralysis, epilepsy, etc., but finds few cases as a result of them.

Dr. Wallin gives a valuable summary of opinions among experts as to the degree of heredity among feeble-mindedness. His study shows that they are quite divergent in their beliefs. It seems that the "facts are not yet all in," and that the causation of feeble-mindedness is as complex as its conditions. It seems that from study (that) we cannot depend solely upon colonization, sterilization and marriage laws for the reduction of feeble-mindedness, but we must also prevent toxication of the parents as that is a big factor in injuring the young child.

Ohio University M. LA VINIA WARNER.

J. W. W. WALLIN. *An Investigation of the Sex Relationship, Marriage, Delinquency and Truancy of Children Assigned to Special Public School Classes.* Miami University, 1922, pp. 34.

This article is a summary of a final report to the Board of Education of the St. Louis Public Schools. After seven years of examining children for Special Education, Dr. Wallin concludes that the Board of Education would have to make provisions for more mentally defective boys than for mentally defective girls. He does not give any reasons for this. He found among the Special Class enrollment an average of 2.14 per cent per family and a per cent of 3.3 of all the children enrolled in Special Education to have a record of delinquency. He makes a valuable discovery relating to the criminality and vicousness of the feeble-minded. His discoveries are contrary to previous beliefs in this matter. Dr. Wallin suggests that the Special Class training given these children has corrected this anti-social conduct. The same thing is true of truancy. Dr. Wallin found only 4.6 per cent of the total enrollment of the children in Special Education to be rated as truants. The entire article is well-rounded argument in favor of Special Education.

Dr. Wallin gives a further study of the industrial records of children in Public School Special Classes for mental defectives and finds that the records both industrial and criminal so far as they go are quite satisfactory and point to the value of Special Education. However, a study of the occupations entered show that the Special Schools have given very little specific training in preparation for them and the employment records of these children furnish little evidence as to the type of vocational training which should be supplied in the Special Schools.

The remainder of the article is taken up in a discussion of the Legislative Act in the interest of Special Education for the State of Missouri.

Ohio University M. LA VINNA WARNER.

STRONG AND UHRBROCK. *Job Analysis and The Curriculum.* Personnel Research Series. Williams and Wilkins Company. Baltimore, 1923. Pp. 146. Price $1.00 paper, $2.00 cloth.

There is no doubt that what we need chiefly in our attempt to apply psychology to industrial management, is a large mass of ordered material collected as a result of an actual study from field experiment in problems of industrial management. Be it problems of employment, fatigue or that of incentive, the important thing is not so much the ideal conditions as the creation of the possibility for experiment and in that way learn, in the long run, what can be done in order to reach some workable scheme of human relations in this or that particular industry. Such studies have been published by Link, Chapman, Myers Scott Company, Carnegie institute of Technology and others—to our delight a new study of an important industrial problem by Strong and Uhrbrock has just appeared.

Messrs. Strong and Uhrbrock set before themselves the task of the actual field study of the functions of the executives in the commercial printing industry. This survey "was made for the express purpose of inventorying the duties of executives and discovering where the emphasis should be placed in training men for similar positions" (p. 71.) "Not only was it worth while to secure more comprehensive information upon which to decide differences of opinion between the printing department faculty Carnegie Institute of Technology and printing executives, but also in order to decide differences of opinion between the faculty in the printing department and other departments in the Carnegie institute" (p. 71.) To supply material from which a well balanced and proper curriculm could be developed and organized the authors set before themselves the task of making a comprehensive study of the various executive jobs within the printing industry—the first task was to make a job analysis and from that to develop the job specification. These would naturally supply the information necessary for the empirical curriculm making.

The present study, however, is more than an attempt to record the results of the above stated survey; as a matter of fact it is also an attempt to define the meaning of such an experimental attempt and to state the technique of obtaining the necessary data and result. Consequently we find in the book chapters dealing with such subjects as the following: job analysis, job specification, the construction of a curriculm based on job analysis, and a bibliography of job analysis material.

Some "industrial psychologists" are coming to the conclusion that a job analysis, made for the purpose of deriving a job specification—to be used by the employment and training department and by the wage committee—should contain a study of the actual functions of a job rather than of a time and motion analysis of the worker. This latter type of job analysis has its value, but in a lifferent realm, which is very important—in a study of efficient vorking methods. A job analysis is perhaps not the best name ecause of its traditional connotation to designate a study of the lifferent functions performed by any given worker. Such study nay perhaps be best designated by the name "occupational description." This name seems to us to be more descriptive of the nature and content of such a study. The authors define a job analysis to be "a method of scientifically dissecting a job in order to determine the component elements and their influence upon the length of learning period of the worker, production and labor turnover" (p. 22.) This definitian we fear is too broad; it covers more than the type of study made by the authors; it includes, it seems to us, even the most minute analysis of the worker's time and motion, and of the method of handling himself, tools and raw màterial.

A job specification, the authors say, is the records of results obtained from the job analysis (p. 23.) "The job analysis is the means; the specification is the end" (p. 24.) "Today the most satisfactory job specification is one that lists the specific duties in outline form and contains a complete analysis of duties, man requirements, working conditions, essential education and promotional possibilities" (p. 27.) While in the actual study itself—and to our admiration—the authors kept clearly in mind the different aims of a job analysis and those of a job specification in their theoretical statements, as above quoted, we must confess the difference is not clearly stated, if not really confused; for example, the statement that a "satisfactory job specification is one that lists the specific duties in outline form and contains a complete analysis of duties. "Is that not a part of job analysis or are the nature and content of a job analysis and of a job specification the same? Or is the difierence only that of function?

The content of the book is a valuable contribution to the much needed *experimental* work in personnel problems. Not only the authors and the psychologists in the Carnegie Institute of Technology but also the editors of the Personnel Research Series should be commended for the type of work that they all made possible to appear. It is only to be regretted that more of such studies have not appeared up-to-date but we look with confidence to the future.

Northwestern University A. J. SNOW.

EDWIN A. SHAW AND EDWARD A. LINCOLN. *A Comparison of the Intelligence and Training of School Children in a Massachusetts Town.* Harvard Monographs in Education, Number 1. Cambridge, Mass. The Graduate School of Education, Harvard University. 1922.

We have in this monograph a report of two surveys of the schools of a small town. Group tests of intelligence and educational tests in arithmetic, writing and reading were given in 1921 and again one year later. The study emphasizes the importance of the interpretation of survey results by the teachers and officers themselves. It is upon them then that the duty of providing remedial measures rests. The reports which are not effective in practice in the systems where the data are gathered fail of their main purpose. The authors also caution against over-emphasis of the formal processes which standard tests now measure; and the opinion is expressed that most standards are too high for practical purposes thus causing a sacrifice of the less easily measurable to the more easily measurable results.

WALTER F. DEARBORN, EDWARD A. LINCOLN AND EDWIN A. SHAW. *Standard Educational Tests in the Elementary Training Schools of Missouri.* Harvard Monographs in Education, Number 3. Cambridge, Mass. The Graduate School of Education, Harvard University. 1922.

This monograph gives the detailed results, with some comment thereon, of the educational tests which were given in the investigation of teacher-training in Missouri. A general summary of the results has already been published in "The Professional Preparation of Teachers for American Schools" issued by the Carnegie Foundation. It seems strange that in this university monograph the name "Cape Girardeau" should be almost consistently misspelled.

Ohio University H. G. GOOD.

A. B. FITT. *The Human Instincts in Business.* The Lothian Book Publishing Company, Melbourne and Sydney, 1922. Pp. VIII and 100.

The author of this book does not claim to have made any new contribution to our knowledge of instincts; he has attempted only to point out in what way our knowledge of human nature—in the major part our knowledge of instincts—can be of value in solving the problem of industrial relations, and that seems to be mainly a problem of incentives. The "book is an outline of some of the main features of a series of lectures delivered in 1918 to a body of men and women representative of the Melbourne business world" (p. VIII).

Professor Fitt in his attempt to throw some light upon the difficult industrial problem shows a pretty thoroughgoing influence of Professor McDougal—while his application of Professor McDougal's general theory of instinct is clear it lacks the brilliancy of Professor Watts' treatment in his *Introduction to the Psychological Problems of Industry.* One is familiar with such an attempt i. e., as started in the United States by Professor Parker and followed with last mastery by Mr. Tead; in England by Professor Watts. Neither Mr. Tead's nor Professor Fitt's books are of great value as proposed solutions of the problem of economic motives—their worth lies in a different direction. Mr. Tead's vivid examples of industrial maladjustment are useful as a source of cases; while Professor Fitt's application of instincts to the general problem of economic motives is suggestive. Both, however, suffer because of a lack of method of approach to the problem. It seems to us that the only method is an empirical one and no other. The tendency today is to base too many deductions and generalizations of wide economic application on our hypothetical knowledge of instincts. It is a pity that we should spend so much of our energy in rationalization. If we had knowledge of social behavior which was derived from experimentation or observation of a very large number of social phenomena then with caution we might apply such knowledge to our general problem of economic motives, but as the case is our present social psychology is the most hypothetical branch of the science and therefore such a procedure is at best only suggestive and at times—if taken seriously—dangerous. What we might do, however, is to collect a large number of cases and from these, if possible, infer certain hypothesis, and further by their application—social and economic experiments—come closer to a true knowledge of these complex human motives. If not, the result of such a set of lectures as these by Professor Fitt will produce comments such as the one stated in the introduction by the sponsor of the series: "while the hearer could not say that he had gained anything that he could directly apply in his business, he found that his point of view had been changed" (p. VII).

Just as Mr. Tead's *Instincts in Industry* so Professor Fitt's book *The Human Instincts in Business* suggests—even though unconsciously, I presume, because such a point of view is impossible—that instincts as motives operate one at a time or if in groups as independent units and naturally we find the customary list of chapters, of which the following are some examples: "The Gregarious Instinct," "Suggestion and Imitation," "Self-display, Pugnacity, Contra-suggestion, Emulation," "The Collecting, Acquiring, or Hoarding Instinct," "The Instinct of Self-abasement in Relation to Self-assertion." The following will serve as a specific illustration of the author's application of the supposed knowledge of instincts to a particular problem of industrial management: "Payment by piece rate is the method employed in very many factories (in some countries industrial legislation will not allow it), and there can be no doubt that it gives the most direct approach to the original nature of the instinct. For each separate stroke or effort, for each unit of work completed, there is a separate unit in payment. The reward stands in direct proportion to the work done. The various other methods of task-payment increase the pay with increases in the amount of work done, but the increase is not directly proportional. Still the appeal is greater than that of the fixed wage, which does not present the continual draw or stimulus to effort."

"Again, once the rhythmic activity from effort to reward, reward to further effort is set up, it can be kept going to almost any extent, just as the squirrel or hamster adds to its store long after the winter's needs have been satisfied. Thus, subject to the limits of human capacity, the appeal to effort of the task-payment systems never ceases" (pp. 64-65). Here the author forgets the great dangers of a "piece rate system" upon the general health and efficiency of the worker; the nervous strain upon the slower employee and, what is more, the actual experimental proof that the use of wages as an incentive is rather limited.

It is interesting to note, however, that the British psychologists more than the American—in the United States, however, a group of brilliant younger economists—have attempted to throw some light upon the much vexed problem of economic incentive—a problem so disturbing to all students of industrial personnel.

Northwestern University. A. J. Snow.

University of Iowa Studies in Psychology, No. VIII, Edited by Carl E. Seashore. Psychological Review Co., Princeton, N. J., 1922. Pp. IV, 382. Price (?)

This collection of monographs represents a somewhat extended series of studies in sound localization, the psychology of music, and muscular reaction. It will be impossible to review all the 17 articles in any thoroughgoing way because of their variability of subject matter, and variety of technique, since nine of the articles are doctors' theses.

The first three articles by Seashore, Halverson, and Stewart, respectively, deal with the wave-phase localization of sounds. Seashore sketches the progress made in the subject previous to the World War; Halverson presents a mass of observed data on the subject, but sets forth no conclusions; while Stewart, approaching the problem from the standpoint of physics, reaches the conclusion that the phase effect is not wholly due to varying intensity-ratios but that the organs of hearing respond to the phase as such.

The next eight articles deal with the psychology of music. These are more or less condensed doctors' theses. Bunch investigated the acuity of hearing throughout the tonal range. He reviews the previous work done in this field, describes the Iowa Pitch Range Audiometer which he perfected, and presents data to show wherein his instrument permits a much more effective means of detecting various types of defects of the internal ear than any instruments in previous use. Zuehl's thesis describes a still further development of the instrument described by Bunch. He also establishes norms of auditory acuity based upon ages of 6 to 15, 17 to 41, and 42 to 73 years, in which some very interesting facts with regard to pitch acuity for high tones is shown. Zuehl presents a very brief article describing a stroboscopic device for measuring revolution rates. As a piece of laboratory equipment this apparatus meets a difficult situation at reasonable cost. Knock has investigated a means of training the voice by the aid of the eye. His measurements are made with the tonoscope, the standard 256 fork, and the Koenig resonator. His subjects sang a standard tone, its major third, its fifth, and the octave. His results agree with previous work in that the major third, the fifth, and the octave are found to be of about equal difficulty in singing. Errors in singing tend to a daily similarity. It is found that one can judge the pitch and inaccuracies of another voice better than of his own. Gaw applies the principle of vocational guidance to schools of music. Her data collected from students in music at Northwestern University shows clearly by graphic representation of individual characteristics that musical surveys, thoroughly made, are of great value in advising students who desire a musical education. Stanton's study of inheritance of specific musical capacities presents family histories, and a series of musical talent pedigree charts which lead her to the conclusion that musical talent is probably inherited according to Mendelian laws. Merry's article describes a method of graphically recording all the speech inflections of the human voice. A detailed description of the apparatus, including an illustration, is given alongside sample charts and graphs. Schoen's article is a study of the unconscious modifications of the singer's voice while attempting to sing a musical melody. Especial attention is given the vibrato. He concludes that tones are usually attacked at a pitch lower than intended when preceded by a lower tone, and that they are released at a pitch higher than intended; that tones are rarely sustained their full interval; that two tones are seldom sung alike even by the same individual that certain vowels are inherently higher than others; and that the emotion expressed by the vibrato is made possible because we have so frequently associated tremor with emotion in past experience. Wickham's article is only a preliminary study in the control of intensity of sounds by touch. Agnew presents two articles in which she attempts to compare the auditory imagery of musicians, psychologists, and children. She uses the questionnaire method and finds that musicians generally attribute themselves with an auditory imagery which develops with training; Psychologists are more critical and believe that their musical images are more kinaesthetic than auditory; and children and untrained adults allege a very decided auditory imagery of sounds. Her study of great composers shows that they invariably possessed a vivid auditory imagery, but this was usually supported by strong imagery in other fields.

The last three articles are studies in motor capacities. The first, by Miss Koerth, describes a pursuit apparatus designed to study eye hand coordinations. A phonograph disc, with stylus, magnets, counter, etc., compose the apparatus. Ream's article summarizes previous work with a tapping test, and describes a new technique for this much used experiment. Hensen's article is a study of serial action. His apparatus consists of a commutator attached to a typewriter in such a way that any one of four circuits may be closed. His subjects were students, army recruits, musicians, and stenographers. He finds a decided personal equation in speed, no consistent relationship between speed and accuracy, some correlation with other motor tests, and slight positive correlation with other motor tests, and slight positive correlation with intelligence tests and ability to read music at sight.

Ohio University. *C. H. Growdon.*

NEW BOOKS AND PAMPHLETS RECEIVED*

Books and pamphlets for review should be sent to James P. Porter, Department of Psychology, Ohio University, Athens, Ohio.

Accredited Secondary Schools in Maryland. Department of the Interior, Washington, D. C. 3pp.

Annual Report of the General Education Board, 1921-22. General Education Board, 61 Broadway, New York. 86 pp.

Atti Dell' Ateneo di Science, Letter ed Arti in Bergamo. Vol. XXVI, 1921. Istituto Italiano De'Arti Grafiche.

BAWDEN, WILLIAM T. *Studies About Occupations in Public Schools.* Department of the Interior Bulletin, March 1923. Washington, D. C. 34 pp.

BRIERLEY, SUSAN S. *An Introduction to Psychology.* Dodd, Mead & Company, 443 Fourth Ave., New York. Price $2.00. 150 pp.

DEFFENBAUGH, W. S. *Secondary Education in 1921 and 1922.* Department of the Interior, Bulletin No. 12, 1923. Washington, D. C. 30 pp.

DEFFENBAUGH, W. S. *Significant Movements in City School Systems.* Department of the Interior, Bulletin No. 8, 1923. Washington, D. C. 28 pp.

Department of Educational Investigation and Measurement. Boston Public Schools. Boston, Mass. 9 pp.

DUMAS, GEORGE. *Traits de psychologie.* Librairie Felix Alcan, 108, Boulevard, Saint-Germain. 964 pp.

Government Publications of Interest to Home Economics Teachers and Students. Department of the Interior, Home Economics Circular No. 5. Washington, D. C., 16 pp.

GREIG, J. Y. T. *The Psychology of Laughter and Comedy.* Dodd, Mead & Company, 443 Fourth Ave., New York. Price $4.00. 300 pp.

HALBERT, L. A. *What is Professional Social Work.* The Survey, 112 East 19th Street, New York City. 149 pp.

KOHS, S. C. *Intelligence Measurement.* The Macmillan Company, New York City. 312 pp.

LEFEBURE, VICTOR. *The Riddle of the Rhine.* E. P. Dutton & Company, 681 Fifth Avenue, New York City. 274 pp.

List of References on Visual Education. Department of the Interior, Library Leaflet No. 18, March 1923. Washington, D. C. 11 pp.

Lombard, Ellen C. Recent Development of Parent-Teacher Associations. Department of the Interior, Bulletin No. 5, 1923. Washington, D. C. 14 pp.

MEAD, ARTHUR RAYMOND. *Learning and Teaching.* J. B. Lippincott Company, Philadelphia, Pa. 277 pp.

MOORE, HARRY H. *Status of Certain Social Studies in High Schools.* Department of the Interior, Bulletin No. 45, 1922. Washington, D. C. 21 pp.

Preparation of School Grounds for Play Fields and Athletic Events. Department of the Interior, Physical Education Series No. 1, 1923. Washington, D. C. 17 pp.

Publications Available April, 1923. Department of the Interior, Washington, D. C. 20 pp.

Public School System of Arkansas. Department of the Interior, Bulletin No. 10, 1923. Washington, D. C. 79 pp.

---

*Mention here does not preclude further comment.

NEW BOOKS AND PAMPHLETS RECEIVED

ROBERTS, LYDIA J. *The Child Health School.* Department of the Interior, School Health Studies No. 2. Washington, D. C. 60 pp.

Report of the Commissioner of Education. Department of the Interior. Washington, D. C. 32 pp.

Report of the First Commercial Education Dinner Conference. Department of the Interior, Commercial Education Leaflet No. 3. Washington, D. C. 16 pp.

Report of a Survey of the University of Arizona. Department of the Interior, Bulletin No. 36, 1922. Washington, D. C. 89 pp.

SPALDING, K. J. *Desire and Reason.* E. P. Dutton & Company, New York. Price $3.50. 220 pp.

VANDEWALKER, NINA C. *Some Experiments in Preschool Education.* Department of the Interior, Kindergarten Circular No. 10. Washington, D. C. 4 pp.

WHITNEY, ANNE. *Who's Who in Healthland.* School Health Studies No. 3, Department of the Interior. Washington, D. C. 32 pp.

THE
JOURNAL OF APPLIED PSYCHOLOGY
VOL. VII. DECEMBER, 1923. No. 4

EXPERIMENTS ON ROD-DIVINING

BY WILLIAM S. FOSTER
UNIVERSITY OF MINNESOTA

Some months ago the pastor of a church in a neighboring town, a man about 70 years of age, called at the Psychological Laboratory of the University of Minnesota and asked that experiments be conducted to measure, and if possible to explain, his apparent powers in locating materials by means of divining-rods.

He gave an account covering 45 years of successful location, not only of water but of oil, natural gas, iron, gold and silver in Wisconsin, Minnesota, West Virginia and Texas, and he presented documentary statements by engineers and others who had witnessed his successes, amply sufficient in amount and respectability to warrant the belief that his powers are typical, and at least as great and as convincing to observers as are those of diviners in general. His utter sincerity was manifest not only by his profession and by his action in seeking scientific test and explanation, but also through the fact that he agreed without reserve that the time and all other conditions of experimentation should be determined by us, and that publication should be allowed, whatever the favorable or unfavorable outcome of results. Further evidence appears in the account below.

The experiments were conducted in and about the Psychological Laboratory during the morning and afternoon of March 26, 1923. The writer was assisted by Professors R. M. Elliott and D. G. Paterson.

TEST 1. DIVINATION OF LOCALITY

The diviner reported previous uniform success in locating hidden coins and watches, and our first experiment is intended to test this ability. "Neutral ground" was discovered in the main laboratory and a large table moved upon it. The table top was then chalked in 14 squares, each 20 x 20 inches. The center of each square was numbered visibly, beginning with numbers 1 and 2 at the left end and continuing with odd numbers on one side and even numbers on the other. The floor directly under the table was marked in squares exactly to correspond to the top, and the floor and space beneath were shielded from sight by cardboard screens.

The rod used was a piece of fairly stiff piano wire bent in the shape of a small written "e". It was therefore smaller and more flexible and smooth than the Y of sapling often chosen. The "e" shaped loop in the wire was approximately 2 inches long and the ends, which were held between the fingers in a common divining position, were each about 5½ inches in length.

Silver coins, a silver watch and 2 gold watches were placed in a small cardboard box. An experimenter first placed this box upon the floor in the center of a square chosen by the diviner. When the diviner stepped upon the corresponding square of the table top the rod rotated definitely and strongly. When he stood in any other square he perceived not even a tendency to rotation.

A random series of trials was then made. Their order was known only to the experimenter who did the placing. Before each trial this experimenter left the room. The diviner was then brought in by a second experimenter, mounted the table, and walked slowly over the squares until the rod rotated. In each trial he then continued exploring without indication over all the non-critical squares in order, and finally "double checked" his first location. No hesitation was manifested at any time, and often the diviner carried on a conversation as he explored.

The second experimenter recorded the estimates, which always fell in the center of some square. Thirty-two trials were made, the box being placed twice in the center of each of the 14 squares, and at intervals (without the diviner's knowledge of such intention) 4 trials were made in which the box was not placed at all, but removed entirely from the room.

In all of these four latter trials indications were obtained exactly as usual, squares 13, 3, 1 and 5 being designated. In the remaining twenty-eight trials the correct square was indicated once only, whereas two correct judgments are likely to be given by chance alone.

Considering the two "sides" of the table merely (i.e., odd numbered *versus* even numbered squares) the side was correctly indicated in 57% of trials and the wrong side in 43%, whereas 50% of correct judgments are most likely by chance alone. But if we consider the accuracy in judgments of "end" (i.e., first 7 *versus* last 7 squares; and first six plus square No. 8 *versus* first six plus square No. 7) the outcome is more than equally unfavorable; 37% right *versus* 63% wrong.

A consideration of the *amount* of error made in location verifies the judgment that successful locations were rather fewer than chance would indicate. The average error which would most likely occur by chance is approximately 50 inches. *The actual* errors average 54 inches. Another fair (but possibly unreliable) method of judging accuracy is by a "pseudo-correlation" between the square-numbers reported and the numbers given. By rank differences the "correlation" so obtained is −.22. A rank-differences correlation which considers merely the seven cross-table oblongs of two squares each comes out −.13.

We must conclude that under such rigidly controlled conditions *no power of accurate location whatever is indicated.* The assurance and lack of hesitation with which such faulty locations were made by the diviner in the apparently total absence of cues are truly astonishing, but our observation furnished no basis for even a guess at explanation.

TEST 2. DIVINATION OF AMOUNT OF MATERIALS

Neither the diviner nor any experimenter was permitted to know the actual results of the first test, but a second test was begun at once. In preliminary trials (made with knowledge on *his part concerning the amounts of material placed in the box)* the diviner showed that with no materials, the rod did not move; with only the coins, the rod moved weakly and slowly; with the addition of one watch it moved perceptibly more strongly and faster; with still another watch yet more; and with another watch added, very perceptibly more. He maintained his ability to discriminate with ease between these steps in amount of material present, and agreed to report accordingly in an experiment either "nothing," "weak," "fair," strong," or "very strong."

The diviner stood on a square on the table top. An experimenter sat beneath the table and placed in random order in the box in the center of the corresponding square on the floor either nothing or an indicated amount of material. Twenty-five trials were made; 5 with nothing, and 5 with each of the four positive amounts.

The diviner judged correctly in 24% of cases, 20% being indicated by chance. His average error was 1.4 steps, as compared *with the 1.6 steps which chance would indicate.* The *rank-differences* correlation between actual and estimated amounts was +.20, P. E. = ±13, (Pearson coefficient = +.12), as compared with the .00 correlation most indicated by chance alone.

It is impossible to say whether these relatively small differences from chance are significant or not. To avoid the fatigue of constant mounting on the table, the diviner did not leave the room between trials, and the possibility exists that differences in the noises made in changing the amounts, in placing the box again on the floor, in the tone of the experimenter's voice in calling "ready", etc., may have served quite "unintentionally" or "unconsciously" as cues which slightly favored correct judgments.

TEST 3. DIVINATION OF DEPTH

As an indicator of depth, the diviner used a heavy straight wire rod about 20 inches long, fitted with a handle and weighted at the outer end. This rod vibrates vertically a given number of times until it corresponds to the supposed depth of the material in feet below the diviner's ground. The plane of vibration then suddenly shifts to horizontal.

"Neutral ground" was located in the hallway on the second floor of the laboratory. In the lower hall directly beneath, a long step-ladder reaching from floor to ceiling was set up at an angle of approximately 70 degrees. The depths chosen for experiment were the 12 steps of this ladder and a point on the floor at its foot. According to measurements taken after the completion of the experiment the rise per step was approximately 11 inches, and the range of the thirteen depths chosen varied between approximately 1.1 and 11.9 feet.

The experimenter placed the box of materials used in test 1 at various depths in a random order known only to himself. He was never visible to the diviner, but noises made in climbing up and down and in placing the box were faintly audible *via* the stairways to any one in the second floor hallway where the diviner stood. The possibility of such auditory cues might have been avoided by leading the diviner from the hall between trials but time did not permit. The diviner and the second experimenter, who stood by him and recorded his estimates, knew nothing of the actual depth-positions used save that they varied by steps between the two floors.

Twenty-six trials were taken in random order, each of the 13 depth positions being twice used. In each trial the diviner used first the location rod and then the depth-rod described above. He then gave verbal reports. Estimates such as "between 7 and 8 feet" were recorded as 7½ feet.

The estimates ranged between 3 and 15.5 feet. Seventy-seven per cent of trials were over-estimations, the average estimate being 9.5, as compared with a true average of 6.5 feet. The average error of estimate was 4.5 feet (in a range of about 11 feet) and only 12% of the judgments were accurate within 1 ft. The rank-differences correlation between the actual depths and the diviner's estimates was +.29, (P. E. = .11) which even in so few trials ought perhaps to be regarded as significantly better than chance.

TEST 4. DIVINATION OF DEPTH WITH INTENTIONAL SUGGESTION

About two-thirds of the way up the stairways leading from the first to the second floor hallway mentioned in test 3 are two landings, from which both the step-ladder and the diviner could be seen. In test 4, taken in the afternoon, six spectators were allowed to stand and watch both the placing and the divining. In order clearly to see the experimenter place the box on the upper steps of the ladder, spectators naturally lean forward slightly.

The diviner was told simply that further experiments on depth-estimates were desired to check those of the morning, and that some of our graduate students and instructors who had never seen divining would be interested to observe it. The spectators did not in general know the purpose of the experiment save that a test of divining was being made. They were warned not to talk loudly, nor by smiling when estimates were made to give possible indications of accuracy or inaccuracy. After half of the trials were made the spectators were told privately to try to make their observation of the experimenter's "placings" somewhat less visibly evident by avoiding leaning to see the upper steps of the ladder. All watched with some interest a series of 26 trials, which in all other respects exactly duplicated those of test 3.

The estimates ranged between 2 and 15 feet. Only 58%, instead of the 77% of test 3 were over-estimations. The average estimate was 6.9, as compared with the former average of 9.5, and with the true average of 6.5 feet. The average error of estimate was reduced from 4.5 to 2.3 feet, and the percentage of estimates accurate within 1 foot was raised from 12 to 19. After the twenty-sixth trial, 5 successive additional trials, made as if in continuation of the series, but with the box of materials entirely removed from the hall, resulted in estimates of 2, 14, 3, 7½ and 3 feet.

The fact that after test 3 the diviner had seen the experimenter measure with a yard-stick the total height from floor to floor may partially account for the improvement in the range of his estimates, and for some of the decrease in amount of over-estimation and of errors, though there is no evidence to show that such correction was realized or knowingly made. Such knowledge of absolute distance, however, would not by itself tend to raise a *rank-differences* correlation above that obtained in test 3, since *relative* rank-order (and not absolute accuracy) is here the determining factor. The correlation, however, was actually raised from +.29 to +.77, and this notable improvement in accuracy of relative depth accordingly appears to be due to cues obtained from the spectator's behavior (unconscious and involuntary movements).

Spectators agree that external observation of the diviner's behavior did not at all indicate that he gave any particular or *intentional* attention to them or their actions. When not conversing he appeared simply to be "entrancedly" attentive to the action of the rod. He himself, with every objective indication of sincerity, maintains that the rod simply "acts of itself", independently of or "almost in spite of" his volition, and he has no theory to account for his apparent and real successes under less strictly controlled conditions in the field.

He was entirely surprised when shown the results of the tests described, and maintains that his experience had led him to believe that his estimates under our conditions would be nearly or quite perfect. When our theory as to cues was propounded after experimentation, he still could not believe that such a theory can be applied completely or adequately to explain his previous successes, and he made no attempt to justify or to explain the results here presented.

TEST 5. THE LOCATION OF WATER-MAINS

In order more nearly to approximate field conditions, the diviner was asked to judge, through indications of location, direction, width and depth, the probable course of water-mains leading to two buildings on the Campus. An experimenter, who knew the true locations, accompanied him and recorded the observations on a rough map. Light snow covered the ground in most places.

The results obtained are difficult to describe without reproducing the map, but our general impression is that, as is usual in such cases, the diviner's judgments while by no means perfectly accurate or complete, were far too accurate to attribute solely to chance. In case of the first building the only probable course indicated was near the true main, which leads 250 feet to a neighboring building, but it was distant about 15 feet at one end and about 50 feet at the other from the true course. The indicated depth of 11 feet is approximately the true depth. A second indicated course of about 100 feet at 30 feet depth was at an angle to a sewer-pipe leading from the rear of the building and opening out of sight about 50 feet down the river-bank.

In the case of the second building an L-shaped course of 80 feet at 10 feet depth, ending at a visible hydrant was indicated exactly over a true main, but a second main leading to the building was passed over entirely without indications. These out-of-door observations appear typical and in all essential respects similar to numerous observations reported by Barrett[1] and others. Such experiments point to a judgment that under field conditions, where surface signs and cues from the accompanying person are possible, the diviner is usually more accurate than chance would indicate.

SUMMARY

A fair summary of the evidence regarding the accuracy and the explanation for the apparent successes of diviners is entirely impossible in short space.[2] In general, however, the evidence appears to the writer to justify the following conclusions:

1. The matter cannot be dismissed by saying that the diviner's successes in the field are simply lucky or chance hits, but they probably far exceed the expectation of chance.

2. Diviners are not in general charlatans, but sincere believers in their powers. They do not in general possess accurate geological knowledge, nor do they in general appear to rely greatly upon *voluntary* observation of contours and other surface signs, such as experience may have taught them might be significant. If such signs are important, as they appear to be in some cases, knowledge of their preception and importance, and knowledge of his own reliance upon them is not in general realized by the diviner.

3. In almost all "field experiments" the diviner has been accompanied by a "checker" or person who possessed knowledge of the true locations, depths, etc., and cues given unconsciously by such persons in their involuntary movements or unguarded speech appear to be a common basis for the diviners indications, even though again the diviner may not, and usually does not, *intentionally* observe such signs or realize their importance.

4. Experiments in which the cues mentioned in 3 and 4 above were rigidly excluded are rare, and their results in general are either inconclusive or actually negative. The experiments in which predictions made by geologists and by diviners disagreed, and where wells were later drilled to ascertain the facts, do not appear crucial, and the results are variously estimated and explained by different investigators.[3] Two previous experiments under conditions somewhat similar to our own test 1 are briefly

---

[1] See F. A. Barrett, *Proc. Soc. Psychical Research*, Vol. 13, 1897, pp. 2-280, and Vol. 15, 1900, pp. 130-383.

[2] A fairly full bibliography is given in *The Divining Rod A History of Water-Witching*, U. S. Geol. Survey Pamphlet 416, 1917.

[3] See e. g. W. J. Sollas: *Proc. S. P. R.*, 2, 1884, pp. 73-78, and W. F. Barrett. *op. cit.*, Vol. 15, pp. 111-61.

reported or referred to.[4] The results appear quite uniformly negative, since the diviners obtained approximately the percentage of successes which chance would indicate. Differences in humidity and temperature, etc. (together with hyperaesthesia) have been proposed as cues without, it appears, great warrant. In view of the facts that diviners appear more successful and believe more in their ability to locate water than metals, running rather than standing, and shallow rather than deep water; that plumbers commonly listen for leaks and that everyone has noticed the peculiar auditory quality of the moving water current, it is strange that one finds on casual reading no suggestion that *sound* may be an important cue in many cases. I am told that in some countries putting an ear to the ground is a recognized method of locating underground springs or courses. Our own experiments certainly point to its possible and probable significance under certain conditions. The psychologist certainly does not, at any rate until further inexplicable facts appear, need to suppose a special sense nor a special mystical power of "clairvoyance" or "telepathy" to explain water-divining.

5. The movements of the rod appear to the diviner to be automatic, and often not even to be subject to his voluntary control. The form of the rod and the mechanics of its movement are such as to favor the suggestion that it "moves of itself" and sometimes moves in spite of voluntary efforts genuinely made to restrain it.[5] From unbelievers who inwardly vow in advance to hold the rod steady "though an ocean should be flowing beneath their feet," reports are not uncommon that they do not succeed, but that the rod "twists the more, the more they endeavor to restrain it." The various forms given to the rod, the success of some diviners who use no rod at all but go by "impressions" or "shiver", and the facts that only a relatively small percentage of persons are "sensitive", that the indications appear only if the diviner is "set" for them, and other facts too numerous to mention, all indicate that the existence of a psychical automatism and the presence of some suggestion to touch off a recognized response are the only necessary and essential bases for the "indications".[6]

This hypothesis is all the more probable in the light of our knowledge of the part played by hyperaesthesia and "unconscious notice" preceding or during trance, and in hypnotic recall, the determination of dreams, automatic writing and the like, and in the light of our knowledge of the importance of cues and suggestions often given by the knower's unconscious and involuntary actions and accepted by the "sensitive" without clear realization of the fact. To become convinced in the latter regard, one has only to compare the conditions and results of divining with those of mind-reading, muscle-reading, table-turning and of thinking animals like Clever Hans.

---

[4] See E. R. Pease: *Proc. S. P. R.*, 2, 1884, pp. 70-94, and W. F. Barrett, op. cit., Vol. 13, p. 176; Vol. 15, p. 360.

[5] See, for example, numerous instances given by W. F. Barrett: *op. cit.*, M. Culpin: *Spiritualism and the New Psychology*, 1920, pp. 34-43.

[6] "Accompanying the involuntary and unconscious muscular contraction which causes the motion of the forked twig or rod, many diviners experience a peculiar *malaise* and some a convulsive spasm. * * * The state of monoideism of the douser creates a condition of partial catalepsy when some suggestion causes the idea to culminate." Barrett: *op. cit.*, Vol. 15, p. 314. See also Vol. 13, pp. 272-277, 245-249; Vol. 15, pp. 130-131, and elsewhere.

A STUDY OF THE APPROPRIATENESS OF TYPE FACES

BY A. T. POFFENBERGER
COLUMBIA UNIVERSITY

AND R. B. FRANKEN
NEW YORK UNIVERSITY

In the search for perfect advertising no factor is considered of too slight importance to deserve careful investigation. The laws of human behavior as discovered by the psychologists have been generally called upon in order to draw away from an individual the control of his behavior and put it into the hands of the advertiser. By far the greatest amount of consideration has been given to the mechanical devices for gaining and holding attention, and for making impressions permanent. The feelings, pleasantly and unpleasantly toned, are coming to be looked upon as forces influencing reactions to advertising sufficiently to deserve study. Freudian psychology, with its well-known doctrine of the suppression of the unpleasant, may have something to do with this spread of interest. However, Hollingworth some years ago showed the bearing of the "obliviscence of the disagreeable" upon advertising, and pointed out that advertisements to be most effective should be pleasantly toned. No labor or expense is spared on the makeup of an advertisement and especially in color and art work in order to make it pleasing. Much of this elaborate use of artistic material is sometimes criticized on the ground of lack of appropriateness, for after all is said, it is the appropriateness of the presentation that determines whether or not the total effect shall be pleasing.

Typography has long been believed to be a source of mildly pleasant or unpleasant feeling tone, according to its character and appropriateness, and this belief has influenced the choice of type in advertising. The accompanying cut (Fig. 1) will illustrate the differences in the type faces used for various kinds of commodities. In the case of "Disston" and "Speed-grits" the type very clearly carries something of the atmosphere of the commodity. In the case of the other names the appropriateness may not be so evident. The belief is fairly general that heavy faced type carry with them the atmosphere or feeling of solidity and strength, and that the thin faced type suggest fineness and delicacy. Figure 2 shows in a striking way the type faces that are supposed to

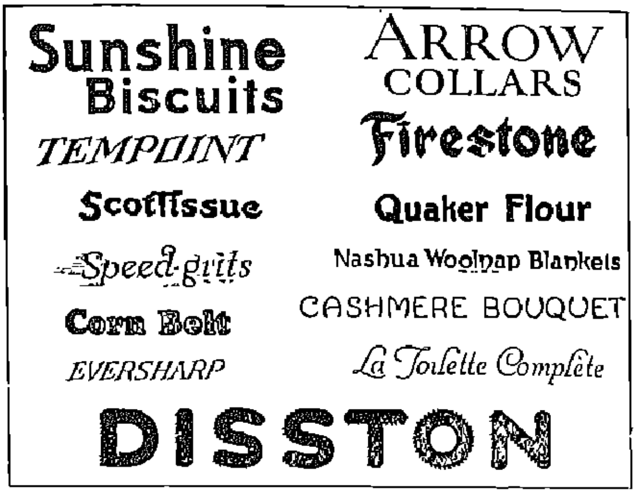


FIGURE 1.

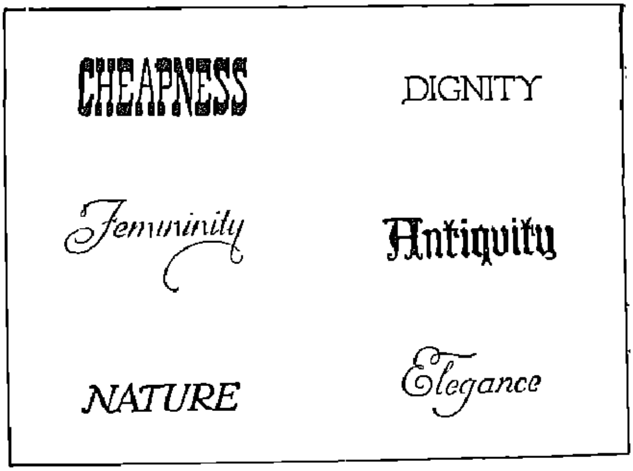


FIGURE 2.

carry the atmosphere of cheapness, dignity, femininity, antiquity, nature, and elegance.

The only published experimental study known to the authors of the appropriateness of type faces is that of Berliner.[1] She measured, by the order of merit method, the relative suitability of eighteen styles of handlettered type for advertising each of four different commodities, namely, fish, pork and beans, pan-cake flour and orange marmalade. Four groups of people were used in the measurements. The coefficients of correlation among the groups for any one commodity will show the degree of agreement among the groups concerning the appropriateness of the type faces for that commodity. The average coefficient of correlation for each of the four articles is given below:

Fish	+.71
Pork and Beans	+.74
Pancake Flour	+.62
Orange Marmalade	+.50

The author interpreted these coefficients as indicating the degree of definiteness of the atmosphere possessed by the commodities, and as the figures show, this definiteness varied somewhat with the commodity.

If now the average orders obtained from any group of subjects for the different commodities are correlated, the size of the coefficients will indicate the degree to which the order of the type varied for the different commodities or what the author called the specificity of the atmosphere. The following figures (Table I) were obtained from the judgments of a group of one hundred people:

TABLE I

	1	2	3	4
1. Fish	—			
2. Pork and Beans	+.75	—		
3. Pancake Flour	+.15	+.20	—	
4. Orange Marm.	−.32	−.15	+.31	—

Thus, the order for fish and orange marmalade tended to be reversed, (−.32), that is, the style of type that was most appropriate for fish tended to be the least appropriate for marmalade. On the other hand, for fish and pork and beans the orders resembled each other very closely (+.75). The author concludes that commodities might be grouped into classes according to the character of their atmosphere and that type might then be chosen which would be appropriate for any member of the class.

---

[1] Ztsch. f. Angewandte Psychol., Bd. 17, 1920.

The experiments to be described in this report differ from those of Berliner in several respects, namely:

1. Instead of hand lettering, the material consisted of twenty-nine of the type faces commonly used for advertising work.
2. Appropriateness of type faces for both abstract qualities and for actual commodities was measured.
3. The judgments of men and women were kept separate in order to show whether sex differences in this form of reaction were present.
4. In only a few cases was the same person asked to judge more than one commodity, and then only after an interval long enough to minimize the effect of the memory of previous arrangements.

A modified form of the order of merit method was employed in this work on account of the large number of specimens to be judged. The type faces each mounted on a separate card were first sorted by each judge into five piles, the first pile containing the most appropriate and the last one containing the least appropriate specimens. The order of the specimens was next determined within each of the five groups. Finally, the specimens in the five groups were put together into a single order. From forty to fifty judgments were obtained in this manner for each of the ten categories given below:

Cheapness
Dignity
Economy
Luxury
Strength

Automobiles
Building Material
Coffee
Jewelry
Perfume

On the following pages are given the samples of type faces used in the experiments with their technical names and also the final rank[1] obtained by each specimen when judged for its appropriateness for each of the categories listed above. Position I indicates the most appropriate specimen, and 29 the least. As there were twenty-nine specimens, the position of any specimen may vary from 1 to 29. The type as illustrated in this report are somewhat smaller than those used in the actual experiment, having been reduced in order to fit the printed page.

---

[1] The figures represent ranks obtained from men subjects only. As will be demonstrated later, the reactions of the sexes in this study are so similar as to make presentation of both sets of figures unnecessary.

A WHEN, in the course of human
events, it becomes necessary $1234&

A—Della Robbia

Cheapness	17	Automobiles	12
Dignity	14	Building Material	16
Economy	5	Coffee	16
Luxury	12	Jewelry	12
Strength	16	Perfume	6

**B WHEN, IN THE COURSE OF HU-
man events, it becomes necessary
for one people to dissolve $12345&**

B—Bodoni Bold

Cheapness	7	Automobiles	6
Dignity	17	Building Material	9
Economy	9	Coffee	10
Luxury	16	Jewelry	20
Strength	9	Perfume	21

**C WHEN, IN THE COURSE OF $12
human events, it becomes 34**

C—John Hancock

Cheapness	2	Automobiles	8,5
Dignity	23	Building Material	3.5
Economy	14	Coffee	2
Luxury	25	Jewelry	25
Strength	3	Perfume	28

D WHEN, IN THE COURSE OF
human events, it becomes $12&

D—Roycroft Tinted

Cheapness	25	Automobiles	22
Dignity	6	Building Material	20
Economy	27	Coffee	18
Luxury	7	Jewelry	7
Strength	19	Perfume	8

E *WHEN, IN THE COURSE OF*
human events, it becomes neces $12

E—New Caslon Italic

Cheapness	15	Automobiles	7
Dignity	12	Building Material	19
Economy	19	Coffee	19
Luxury	11	Jewelry	19
Strength	17	Perfume	10

F WHEN IN THE COURSE OF HUMAN
EVENTS IT BECOMES NEC $1234

F—Blair

Cheapness	21	Automobiles	23
Dignity	7	Building Material	22
Economy	11	Coffee	22
Luxury	10	Jewelry	8
Strength	21	Perfume	11

G

WHEN, IN THE COURSE OF
human events, it becomes $12345&

G—Caslon Old Style No. 471

Cheapness	13.5	Automobiles	15
Dignity	13	Building Material	16
Economy	6	Coffee	11
Luxury	13	Jewelry	13
Strength	18	Perfume	14

H

WHEN, IN THE COURSE OF
human events, it becomes necessary for one
people to dissolve the political $1234567&

H—Caslon Old Style No. 471 Italic

Cheapness	24	Automobiles	17
Dignity	8.5	Building Material	25
Economy	24	Coffee	27
Luxury	6	Jewelry	3.5
Strength	27	Perfume	2

WHEN, IN THE COURSE OF HUMAN
events, it becomes necessary for $123

I—New York Gothic

Cheapness	13.5	Automobiles	20
Dignity	15	Building Material	8
Economy	4	Coffee	14
Luxury	17	Jewelry	22
Strength	8	Perfume	22

J

WHEN, in the course of human events, it becomes necessary for one people to dissolve the political bands which have con $123456&

J—Century Old Style

Cheapness	11	Automobiles	24
Dignity	22	Building material	25
Economy	2	Coffee	18
Luxury	18	Jewelry	21
Strength	24	Perfume	15

K When, in the course of human $1

K—Old English

Cheapness	22	Automobiles	25
Dignity	4	Building Material	23
Economy	28	Coffee	23
Luxury	4	Jewelry	11
Strength	15	Perfume	16

L WHEN, IN THE COURSE OF
human events, it becomes $12345&

L—Goudy Old Style

Cheapness	16	Automobiles	16
Dignity	11	Building Material	17
Economy	8	Coffee	17
Luxury	15	Jewelry	10
Strength	20	Perfume	9

WHEN, in the course of human
M events, it becomes necessary for
one people to dissolve $1234567&

M—Scotch Roman

Cheapness	19	Automobiles	13
Dignity	19	Building Material	13
Economy	3	Coffee	12
Luxury	20	Jewelry	16
Strength	14	Perfume	17

N WHEN, IN THE COURSE OF HUMAN EVENTS, IT BECOMES
necessary for one people to dissolve the political bands which have $1234

N—Circular Gothic No. 44

Cheapness	20	Automobiles	28
Dignity	10	Building Material	26
Economy	16	Coffee	28
Luxury	8	Jewelry	5
Strength	26	Perfume	3

WHEN, IN THE COURSE
O **of human events, it becomes**
necessary for one $1234&

O—Antique Bold

Cheapness	1	Automobiles	3
Dignity	27	Building Material	1
Economy	18	Coffee	3
Luxury	26	Jewelry	28
Strength	2	Perfume	25

WHEN, IN THE COURSE
P **of human events, it $123&**

P—Masterman

Cheapness	5	Automobiles	2
Dignity	26	Building Material	2
Economy	20	Coffee	6
Luxury	23	Jewelry	20
Strength	4	Perfume	24

When in the Course of Human Events, it becomes necessary for
Q *one people to dissolve the political bands which have connected*
them with another, and to assume, among the powers of the
earth, the separate and equal station to which $1234567890&

Q—Typo Slope

Cheapness	28	Automobiles	20
Dignity	5	Building Material	29
Economy	26	Coffee	29
Luxury	3	Jewelry	2
Strength	29	Perfume	1

R
WHEN, IN THE COURSE
of human events, it $123&

R—Century Bold

Cheapness	6	Automobiles	1
Dignity	29	Building Material	3.5
Economy	21	Coffee	7
Luxury	27	Jewelry	24
Strength	5	Perfume	28

S
WHEN, in the course of human
events, it becomes necessary
for one people to dis- $12345&

S—Post Monotone

Cheapness	12	Automobiles	11
Dignity	16	Building Material	11
Economy	7	Coffee	9
Luxury	21	Jewelry	18
Strength	10	Perfume	18

T
WHEN, IN THE COURSE OF
human events, it beco $123456&

T—Caslon Old Style

Cheapness	9	Automobiles	5
Dignity	18	Building Material	12
Economy	10	Coffee	8
Luxury	19	Jewelry	17
Strength	11	Perfume	20

U WHEN, IN THE COURSE OF $12

U—Cheltenham Bold Outline

Cheapness	26	Automobiles	21
Dignity	8.5	Building Material	21
Economy	23	Coffee	20
Luxury	9	Jewelry	6
Strength	23	Perfume	7

V When, in the Course of Human Events,

V—Tiffany Text

Cheapness	29	Automobiles	18
Dignity	2	Building Material	28
Economy	25	Coffee	26
Luxury	1	Jewelry	3.5
Strength	28	Perfume	5

W WHEN, in the course of hu-
man events, it becomes $1234

W—Bookman Old Style

Cheapness	10	Automobiles	10
Dignity	24	Building Material	10
Economy	1	Coffee	13
Luxury	22	Jewelry	21
Strength	12	Perfume	19

X WHEN, IN THE COURSE OF
human events, it becomes $1234&

X—Cheltenham Bold

Cheapness	8	Automobiles	4
Dignity	21	Building Material	6
Economy	12	Coffee	5
Luxury	24	Jewelry	23
Strength	7	Perfume	28

Y WHEN, IN THE COURSE OF H
events, it becomes nec $1234

Y—Globe Gothic Bold

Cheapness	3	Automobiles	14
Diginty	28	Building Material	5
Economy	17	Coffee	1
Luxury	29	Jewelry	29
Strength	1	Perfume	29

Z WHEN, IN THE COURSE OF HUMAN EVENTS

Z—Engraver's Roman

Cheapness	27	Automobiles	27
Dignity	3	Building Material	27
Economy	22	Coffee	24
Luxury	2	Jewelry	1
Strength	25	Perfume	4

X^1 **WHEN, IN THE COURE OF HU**
events, it becomes necessary $12

X^1—Roycroft

Cheapness	4	Automobiles	8.5
Dignity	25	Building Material	7
Economy	15	Coffee	4
Luxury	28	Jewelry	27
Strength	6	Perfume	27

Y^1 WHEN, IN THE COURSE OF HU-
man events, it becomes $123456&

Y^1—Bulfinch

Cheapness	18	Automobiles	19
Dignity	20	Building Material	14
Economy	13	Coffee	15
Luxury	14	Jewelry	14
Strength	14	Perfume	15

Z^1 When, in the Course of Human $123

Z^1—Priory Text

Cheapness	23	Automobiles	26
Dignity	1	Building Material	24
Economy	29	Coffee	25
Luxury	5	Jewelry	9
Strength	22	Perfume	13

An examination of the positions assigned to the specimens when judged for the ten types of appropriateness shows that their relative effectiveness varies according to the purpose for which they are to be used. Differences in the degree of variation are also to be noted. For example, specimen Y^1 varies only from position 13 to position 20 in the ten categories, while Z^1 varies all the way from 1 to 29. The former type specimen would serve moderately well for any purpose, while the latter is either very apropriate or very inappropriate. The same may be said of specimen Z. The range and distribution of the positions assigned to the type specimens may be observed readily from Table II. The positions from 1 to 29 are arranged in six groups, indicated across the top of the table. The figures in the body of the table indicate the number of times that a given specimen fell in a certain group regardless of the commodity for which it was judged. Thus specimen A was twice given a position within the 5-9 group, and 4 times within the 10-14 group. More detailed information concerning range and distribution may be obtained from the figures given under each of the type specimens.

If values be assigned to the groups in Table II, so that the first group is given a value of 1, the second group is given a value of 2, etc., the total value assigned to each type face may be roughly determined. Such total values would indicate which type, if any, tended to be consistently appropriate or inappropriate wherever used. These values are given in Table III. No one of the specimens stands out as generally best or generally poorest. Specimen B appears to be the most effective for all purposes and specimen J the least effective. But the differences are slight.

TABLE III

A	32	G	30	M	34	S	33	Y	36
B	29	H	38	N	41	T	32	Z	38
C	32	I	34	O	32	U	39	X^1	36
D	38	J	42	P	34	V	39	Y^1	36
E	35	K	40	Q	41	W	35	Z^1	41
F	38	L	36	R	35	X	33		

The relation among the orders for the different categories may be expressed in terms of coefficients of correlation, the higher the positive correlation the greater the similarity among the orders for the different categories. The coefficients obtained by the Rank Difference Method are given in Table IV for the men and in Table V for the women.

In interpreting these figures it should be remembered that the

TABLE II

RANGE OF POSITIONS ASSIGNED TO EACH SPECIMEN

Type	Position 1-4	5-9	10-14	15-19	20-24	25-29
A		2	4	4		
B		4	1	2	2	
C	4	1	1		1	3
D		4		2	2	2
E		1	3	6		
F		2	3		5	
G		1	8	1		
H	2	2		1	2	3
I	1	2	2	2	3	
J	1		1	3	4	1
K	2		1	2	3	2
L		2	2	5	1	
M	1		4	4	1	
N	1	2	1	1	1	4
O	4			1		4
P	3	2			3	2
Q	3	1				6
R	2	3			3	2
S		2	4	3	1	
T		3	3	3	1	
U		4			5	1
V	3	1		1		5
W	1		5	1	3	
X	1	4	1		3	1
Y	3	1	1	1		4
Z	4				2	4
X'	2	3		1		4
Y'			5	4	1	
Z'	1	2	1		3	3

groups of persons judging were not the same for any two categories and that rarely was the same judge used twice. Thus, the high negative correlation[3] between the orders for Cheapness and Dignity (–.34) or between Cheapness and Luxury (–.94) represents a real difference in effectiveness and not merely a tendency on the part of the judges to be logical, in placing a given specimen high for one quality and low for some contrasting quality. The coefficients for Cheapness and Economy (+.47) and for Dignity and Luxury (+.96) are high and positive. Strength correlates positively with Cheapness (+.82) and with Economy (+.31) while it correlates negatively with Luxury (–.96) and with Dignity (–.89). Plain heavy type, while indicating strength, also suggest cheapness and economy as compared with the more delicate and ornate type forms.

A coefficient of correlation of +.94 between the orders for Jewelry and Perfume indicates that the type faces that are most appropriate for the one are also most appropriate for the other. These two commodities correlate with the orders for Automobiles and Building Material –.95 and –.92 respectively, showing that the type forms most effective for the one group are the least effective for the other group.

The order for Coffee, while quite similar to that for Building Material (+.90) and for Automobiles (+.81) is quite different from that for Jewelry (–.91) and Perfume (–.96).

Interesting comparisons may be made between the abstract group (Dignity, Economy, etc.) and the commodity group (Automobiles, Building Material, etc.) Cheapness correlates positively and high with Coffee (+.90) and negatively and high with Jewelry (–.94,), and Perfume (–.88). Dignity correlates high and positive with Jewelry (+.89) and with Perfume (+.80). The orders for Economy and Automobiles show a correlation of +.28, Economy and Building Material +.40, and Economy with Coffee +.38. There is a negative relationship between the orders for Economy and Jewelry (–.38) and Economy and Perfume (–.29). The coefficients in which Economy is concerned are the lowest to be found in the table. Luxury correlates high and positively with Jewelry (+.93) and Perfume (+.88) and negatively with the other commodities. Strength correlates high and positively with Automobiles (+.73), Building Material (+.93) and Coffee (+.91), but negatively and high with Jewelry (–1.00) and Perfume (–1.00).

---

[3] The following interpretations will be based on the figures for men given in Table IV.

TABLE IV

CORRELATIONS AMONG THE ORDERS FOR MEN

		1	2	3	4	5	6	7	8	9	10
Cheapness	1	—	-.84	+.47	-.94	+.82	+.78	+.93	+.90	-.94	-.88
Dignity	2	—	—	-.73	+.96	-.89	-.76	-.91	-.85	+.89	+.80
Economy	3	—	—	—	-.48	+.31	+.28	+.40	+.38	-.38	-.29
Luxury	4	—	—	—	—	-.96	-.79	-.94	-.92	+.93	+.88
Strength	5	—	—	—	—	—	+.73	+.93	+.91	-1.0	-1.0
Automobiles	6	—	—	—	—	—	—	+.82	+.81	-.74	-.64
Bldg. Material	7	—	—	—	—	—	—	—	+.90	-.95	-.92
Coffee	8	—	—	—	—	—	—	—	—	-.91	-.96
Jewelry	9	—	—	—	—	—	—	—	—	—	+.94
Perfume	10	—	—	—	—	—	—	—	—	—	—

TABLE V

CORRELATIONS AMONG THE ORDERS FOR WOMEN

		1	2	3	4	5	6	7	8	9	10
Cheapness	1	—	-.91	+.55	-.87	+.94	+.93	+.99	+.80	-.95	-.94
Dignity	2	—	—	-.50	+.90	-.91	-.88	-.91	-.75	+.92	+.92
Economy	3	—	—	—	-.52	+.44	+.41	+.54	+.63	-.49	-.46
Luxury	4	—	—	—	—	-.83	-.81	-.86	-.64	+.88	+.87
Strength	5	—	—	—	—	—	+.92	+.95	+.82	-.93	-.85
Automobiles	6	—	—	—	—	—	—	+.93	+.75	-.94	-.91
Bldg. Material	7	—	—	—	—	—	—	—	+.80	-.95	-.95
Coffee	8	—	—	—	—	—	—	—	—	-.76	-.73
Jewelry	9	—	—	—	—	—	—	—	—	—	+.94
Perfume	10	—	—	—	—	—	—	—	—	—	—

TABLE VI.

CORRELATION BETWEEN MEN AND WOMEN FOR DIFFERENT CATEGORIES

Cheapness	+.91	Automobiles	+.81
Dignity	+.92	Building Material	+.97
Economy	+.93	Coffee	+.79
Luxury	+.93	Jewelry	+.94
Strength	+.94	Perfume	+.95

No differences worth noting appear between the records for men and women. (Tables IV and V). Table VI showing the relationship between the orders for men and women contains no coefficient lower than +.79. This similarity of the reactions in the two sexes gives additional evidence of the consistency of the judgments of appropriateness of material of this sort.

The results of this experiment show quite conclusively that differing type faces do vary in appropriateness and that judges are able to "feel" this appropriateness or lack or appropriateness. Furthermore, there is close agreement between sexes and among members of the same sex in the character of their reactions to the different type specimens.

The experiment produces no measure of the strength of these "feelings". Certainly, the pleasantness or unpleasantness of the effects produced by these different type faces must be extremely mild, and it might be argued that such differences as do exist are too slight to warrant consideration for practical purposes. However, the appropriateness of type is one factor among the many which determine the appropriateness of the advertisement as a whole and therefore cannot be safely neglected.

Some theoretical interest may attach to the question of the cause of the differences found in this experiment. There are two possible explanations. First, the type faces, by their shape, size, texture and the character of their lines may carry a certain atmosphere about them; there may be inherent in them an atmosphere of delicacy, dignity or strength. Second, the atmosphere which they possess may be merely a matter of use. For example, we are accustomed to seeing Old English or similar type, such as specimens K, V, or Z' used for imposing and dignified documents and plain, heavy type such as specimens O and Y for simple and ordinary printing purposes. By the process of association the type may now carry the qualities transferred from the settings in which they have habitually appeared.

For present practical purpose it would seem to make little difference which of these interpretations is correct. It is sufficient to know that type faces such as are used in current advertising practice do have characteristic atmosphere and that these should be given due weight.

One further point should be noted, namely, that the effectiveness of type faces such as we have used in this work cannot necessarily be predicted from the judgment of one individual, even though he may be an expert in typography. Only a test upon an adequate sampling of the population to be influenced will definitely determine what is the best type to use. The judgment of one individual may be safer where the more extreme forms of handlettering are under consideration, although even here some further check would be advisable. The effect upon a sampling of prospects must here, as in other cases of control of human behavior, be the final measure of what will or will not be effective.

INTELLIGENCE TESTS VERSUS ENTRANCE EXAMINATIONS AS A MEANS OF PREDICTING SUCCESS IN COLLEGE

By Ada Hart Arlitt and Margaret Hall
Bryn Mawr College

This investigation begun in 1919 and completed in June, 1923 was undertaken to determine (1) the relative value of the record in intelligence tests and the record in entrance examinations as a means of predicting success in a college which admits only through the regular channels of entrance examinations, (2) the extent to which the record in intelligence tests could be used to exclude students who succeeded in passing entrance examinations but were not capable of doing adequate work in college, (3) the relative merits of two intelligence tests and entrance examination grades as a means of selecting the students who will do the least adequate work in college.

The subjects, 305 in all, were students who entered Bryn Mawr in 1919, 1920 and 1921. The test results for each class were treated separately. As a matter of convenience the classes will hereafter be termed Group I, Group II, and Group III in the order of the date of entrance into college. Students who withdrew for any reason other than poor work before the end of Freshman year and students who were given only one of the two test series in the class to which two sets of tests were given are not included in this study.

Group I was given the Stanford Revision of the Binet Test used in part as a group test. Interpretation of Fables (XVI, 2), Memory for Paragraph (XVIII, 4) and the Vocabulary Test (XVI, 1 and XVIII, 1) were given to the group as a whole. In the first two of these tests the procedure followed was the same as that outlined by Terman except that the responses of the students were written in a regulation college "Blue Book". For the Vocabulary Test each student was given the list of one hundred words printed with sufficient space after each word to allow for the writing of a definition. Ample time was allowed to finish each test. The Vocabulary Test was given last and the paper handed in as soon as completed. All of the other tests were given individually by the writer.

Group II was also given the Stanford Revision of the Binet

The writer wishes to acknowledge her deep indebtedness to Miss Edith Orlady, Secretary and Registrar of the College, for her constant co-operation and courtesy in furnishing the data as to entrance examination records and grades in college.

Test with the same procedure as that described above except that the Ingenuity Test (XVIII,6) was also given as a group test.[1] As soon as the time allowance of five minutes for each section of this test had elapsed the students turned over a page and began the new problem on a new sheet. In order to avoid work on a problem after the time limit for that problem was up, no student was allowed to turn back a single sheet for any purpose. The Thurstone Freshman Test, Series 1920 Number IV was also given to this group. Group III was given the Thurstone Freshman Test, Series 1919.

The results of the tests of Group I in terms of I. Q. were correlated with grades received in entrance examinations and with marks at the end of the Freshman, Sophomore and Junior years and at the middle of Senior year. Marks at the end of the Freshman and Sophomore years are in terms of the average of all percentage grades received by the student during the year. Marks in the Junior and Senior years are in terms of honor points.[2] Grades received in entrance examinations were also correlated with marks at the end of the Freshman, Sophomore and Junior years and at the middle of Senior year. The formula used was the Product Moment with computations from the true average:

$$\frac{\sum xy}{\sqrt{\sum x^2}\ \sqrt{\sum y^2}}$$

The results are presented in Table I.

TABLE I

No. of students		r	P. E.
98	Entrance grades and marks end of Freshman year	.539	.048
83	Entrance grades and marks end of Sophomore year	.548	.052
74	Entrance grades and marks end of Junior year	.541	.055
74	Entrance grades and marks middle of Senior year	.542	.056
98	I. Q.'s and Entrance grades	.245	.064
98	I. Q.'s and marks end of Freshman year	.298	.062
83	I. Q.'s and marks end of Sophomore year	.097	.073
74	I. Q.'s and marks end of Junior year	.102	.078
73	I. Q.'s and marks middle of Senior year	.100	.078

As will be seen from Table I, I. Q.'s show a positive but low correlation[3] with marks at the end of Freshman year and a negligible correlation with marks at the end of Sophomore and Junior years. Grades received in entrance examination show a marked positive correlation with marks received by the end of each of the three college years. The coefficient of correlation for entrance grades and marks received in academic work is almost twice as large as that for I. Q.'s and marks at the end of Freshman year and is more than four times as large as the coefficient of correlation for I. Q.'s and college marks at the end of Sophomore and Junior years and at the middle of Senior year.[4]

---

[1] This test did not lend itself well to use as a group test in as much as the time allowed for each unit was too great for over half of the students tested, but all had to wait until the end of the time allowed.

[2] A certain number of honor points is allowed for each hour of work of High Credit, Credit and Merit grade.

The results obtained from correlating the two intelligence tests used with Group II with marks received in academic work, and entrance examination grades with college marks are presented in Table II. I. Q.'s show a positive but low correlation with college marks at the end of Freshman year (.197±.064) and at the end of Sophomore year (.166±.070). Thurstone Test Scores show a higher correlation with college marks at the end of both years than do I. Q.'s. The correlation between entrance grades and college marks at the end of the Freshman year is .520±.048 and at the end of Sophomore year is .486±.055.

TABLE II

No. of students		r	P.E.
103	Entrance grades and marks end of Freshman year	.520	.048
88	Entrance grades and marks end of Sophomore year	.486	.055
103	Entrance grades and I. Q.	.190	.064
103	I. Q. and Freshman marks	.197	.064
88	I. Q. and Sophomore marks	.166	.070
103	I. Q. and Thurstone Scores	.223	.063
103	Thurstone Scores and Freshman marks	.274	.061
88	Thurstone Scores and Sophomore marks	.293	.066

Here, as with Group I, entrance grades show a much higher correlation with Academic success than do the scores in either of the Intelligence Tests.

The results from Group III are presented in Table III. Entrance grades correlate .638±.039 with marks received in the

---

[3] The classification followed here is Rugg's. Rugg, H. O. Statistical Methods Applied to Education, p. 256. (Houghton Mifflin Co. 1917.)

[4] The marked decrease in the size of the co-efficient beyond the Freshman year may be explained in part by the fact that a number of students low in both grades and scores in intelligence tests withdrew at the end of the Freshman year. This does not, however, account for all of the change.

Freshman year while Thurstone scores show a correlation of .347 ±.058 with Freshman marks.

TABLE III

No. of students		r	P.E.
104	Entrance grades and marks end of Freshman year	.638	.039
104	Thurstone Scores and marks end of Freshman year	.347	.058
104	Thurstone Scores and entrance grades	.298	.060

The results presented in Tables I, II and III show that of the three Measures on which prediction of success in the Freshman year in college might be based, grades in entrance examinations rank first, Thurstone Test scores next and I. Q.'s last. In the case of the last two measures the difference between the sizes of the correlation coefficients in the one group to which both tests were administered in 7.7 points at the end of the Freshman year and 12.7 points at the end of the Sophomore year. No one of the three measures show a correlation which may be regarded as "high" according to Rugg's Classification except in the case of entrance grades and Freshman marks in Group III.

A study was then undertaken to determine whether the records in intelligence tests might have been used to exclude students who, although able to pass the entrance examination, made unusually poor records in college. Each group was divided into quartiles according to rank in college marks. Table IV shows the distribution in quartiles according to rank in college marks of the students at or below the 25th percentile in entrance grades, in Thurstone scores and in I. Q.'s. In the case of Groups I and II a low grade in entrance examinations is more indicative that the student will do less than average work in college than is a low score in intelligence tests as is shown by the fact that a larger per cent of the students ranking below the 25th percentile in entrance grade rank below the 25th percentile and median in college grade. Of the students ranking below the 25 percentile in I. Q. in Groups I and II, 47% rank below the median in college grade at the end of Freshman and Sophomore years. Of those ranking below the 25 percentile in Thurstone scores 60% rank below the median in college grade and of those ranking below the 25 percentile in entrance grade 70% rank below the median in college grade. In Group III the intelligence test and entrance grade select the poorer students almost equally well, but the Thurstone test shows

TABLE IV

DISTRIBUTION IN RANK IN COLLEGE MARKS.

Group	No. students below 25 percentile in Intelligence Quotient	Year	No. below 25 percentile	No. between 25 and 50 percentile	No. between 50 and 75 percentile	No. above 75 percentile
I	24	Freshman	9	6	6	3
I	18	Sophomore	5	3	5	5
II	26	Freshman	6	5	8	7
II	21	Sophomore	3	5	7	0
	No. of students below 25 percentile in Entrance Examination					
I	24	Freshman	10	9	3	2
I	17	Sophomore	7	5	4	1
II	26	Freshman	11	6	7	2
II	21	Sophomore	9	5	4	3
III	26	Freshman	10	9	7	0
	No. below 25 percentile in Thurstone Score					
II	26	Freshman	7	9	7	3
II	22	Sophomore	8	5	6	3
III	26	Freshman	10	11	3	2

itself to be slightly superior to the entrance grade. Of the students ranking below the 25th percentile in Thurstone score 80% rank below median in college marks as against 76% in the case of those ranking below the 25th percentile in entrance grades.

As a final test of the selective capacity of the three measures, a comparison was made of the standing in college of the six students ranking lowest in entrance grades, the six ranking lowest in I. Q., and the six ranking lowest in Thurstone scores in each group. This represents approximately the lowest 6% of the group. Of the six students ranking lowest in I. Q. in Group I three were below the 25th percentile of the class in college grades, two were between the 25th percentile and the median, none between the median and 75th percentile and one above the 75th percentile. Two of the three students below the 25th percentile were dropped at the end of Freshman year because of poor work. None of the six lowest in entrance grades were dropped and only one of the six was below the 25th percentile.

The six students in Group II ranking lowest in Thurstone score were distributed as follows: one was below the 25th perecentile, two were between the 25th percentile and median, two were between the median and 75th percentile and one was above the 75th percentile. Of the six ranking lowest in I. Q. two were dropped at the end of the year because of poor work, two were between the 25th percentile and median and two between the median and 75th percentile. Of the six ranking lowest in entrance grades two were below the 25th percentile, three between the 25th percentile and median and one between the median and 75th percentile. None of the six ranking lowest in entrance grades were dropped because of poor work. Of the six students ranking lowest in Thurstone score in Group III one was dropped because of poor work, and of the six lowest in entrance grade one was dropped.

Although the Thurstone Test is, as has been shown by its correlations, a better measure on which to base prediction of success in college than is the I. Q. and although entrance grades is a better measure on which to base prediction of success in college than is the score in either of the intelligence tests, the I. Q. selects the students who will do the poorest work in college better than does either the score in Thurstone Test or the grade in entrance examinations. A study, however, of the distribution by quartiles in rank in college grade of the six students lowest in intelligence tests will *serve to show that to use the results of such tests to eliminate* students who have passed the entrance examinations would be *manifestly unfair.*

The result of our tests show three things (1) that neither the *Thurstone Test score nor the I. Q. correlated as highly with marks* received in college as did grades received in entrance examinations, (2) that the Intelligence Tests could not with fairness be used to exclude students who had passed the entrance examinations since a large percentage of the students at or below the 25th percentile in Intelligence Test score do adequate college work and less than half of the lowest five per cent in Intelligence Tests are excluded from college for poor work, (3) that the I. Q.'s selected the students whose work proved poorest, better than did the entrance examination grade or the score in the Thurstone Test.

That our results are in no way unique can be seen from a comparision of our results with those of previous investigators in the same field. The following table cited from Caldwell brings together the data from a number of such studies.

The coefficient of correlation between college marks and intelligence tests is shown in the column at the right.

Wissler	.09
Calfee	.23
Rowland and Lowden	.37
Waugh	.41
Bell	.31
Kitson	.44
King and McCrory	.39
Caldwell	.44

To the above table may be added the following:

	Year	r	No.students	Tests used
Caldwell	Sophomores	.17	24	Stanford Revision
	Junior	.69	24	Stanford Revision
Kitson	(Second group)	.20	40	16 Selected Tests
Rogers	Seniors	.42	92	Thorndike Mental Alertness A.
	Seniors	.30	92	Thorndike Mental Alertness K.
	Freshmen	.37	182	Thorndike Mental Alertness A.
	Freshmen	.37	182	Thorndike Mental Alertness K.
	Freshmen	.23	182	Reasoning I Rogers
	Freshmen	.12	182	Reasoning Ia Rog.
Colvin	Freshmen	.5981	103	Brown U. Tests
	Freshmen	.4567	210	Army Alpha

The coefficient of correlation in the studies cited in which tests other than the Stanford Binet were used ranged from .09 to .59 but in most instances varies by not more than ten points from our results with the Thurstone Tests. The only results comparable with ours for the Stanford Binet are Caldwell's. Here the coefficient of correlation for Sophomores was about the same as that in our results for Sophomores in Group II but the coefficient of correlation for Juniors was much larger than ours due perhaps to the difference in the size of the groups.

The fact that in our study scores in intelligence tests and college marks do not show closer agreement may be explained in several ways. In the first place, as has been pointed out earlier in this study, the students dealt with are a highly selected group from which the less able individuals have been for the most part eliminated by failure to pass the entrance examinations. All but seven of the 201 students given the Stanford Binet have I. Q.'s above 100 and the large majority of the students have mental ages above seventeen. As to Thurstone scores the lowest student in Group II scores at the 19th percentile according to the norms for this test. The lowest student in Group III also scores at the 19th percentile. The level of intelligence in the case of practically every student is high enough to warrant adequate work in college if interest, habit of work and the necessary preparation are also present. Factors other than intelligence may therefore play a large part in the student's success.

In the second place very intelligent students have often not been *trained to work to the limit of their capacity in the lower schools.* Their mental level has been high enough to enable them to do fairly satisfactory work with a minimum of effort. These students enter college with inadequate preparation as shown by low grades in entrance examinations and with the habit of working only enough to "get through." If this habit persists through all four college years it may lead to a very intelligent student's receiving poor marks.

In the third place some of the most intelligent students are not interested primarily in college work, and unless pride in academic success is sufficient to make these students put forth a determined effort to obtain high grades most of their energy may find its outlet in extra curriculum activities.

Lastly the low correlation between the two intelligence tests used by us (.223) suggests that the validity of one or perhaps of both as tests which will differentiate between mental levels in adults may not be as great as could be desired. It may well be that we are expecting more of the tests of college students than we have expected of tests of children below fourteen. Our problem here is that of differentiating between abilities at the top of the scale, a far more difficult one than that of separating into groups the whole range of mental ability. It is probable that this problem will require as many years for its solution as elapsed between Binet's first experiments with the thirty tests in his original series and the publication of the Stanford Revision of the Binet Tests in 1915.

Full cooperation of college authorities and experimenters in the field of adult testing will, it is to be hoped, produce tests as valuable to the college as the Stanford Revision of the Binet Test has been to the lower schools.

BIBLIOGRAPHY

1. Bell, C. J. *Mental Tests and College Freshmen.* Jour. Educ. Psychol. 7: 381, 1916.
2. Caldwell, H. H. *Adult tests of the Stanford Revision applied to College students.* Jour. Educ. Psychol., 10: 477, Dec. 1919.
3. Calfee, M. *College Freshmen and Four General Intelligence Tests.* Jour. Educ. Psychol., 4: 223, 1913.
4. Colvin, S. S. *Psychological Tests at Brown University.* School and Society, 10: 27, July 5, 1919.
5. King and McCrory. *Freshmen Tests at the State University of Iowa.* Jour. Educ. Psychol. 9: 32, Jan. 1918.
6. Kitson, H. D. *The Scientific Study of the College Student.* Psychol. Rev. Mon. 23, I, 1917.
7. Rogers, A. L. *Mental Tests As a Means of Selecting and Classifying Students.* Jour. Educ. Psychol., 11: 181, 1920.
8. Rowland, E. and Lowden, G. *Report of Psychological Test at Reed College.* Jour. Exper. Psychol. 1: 211, 1916.
9. Wissler, C. *The Correlation of Mental and Physical Tests.* Psychol. Rev. Mon. Sup. III, 6: 1, 1901.
10. Waugh, Karl. *A New Diagnosis of the College Student.* New York Times Mag. Sup. Jan. 2, 1916.

A RE-INTERPRETATION OF THE STATISTICAL METHOD OF ARMY TRADE TESTS

By Arthur W. Kornhauser
University of Chicago

The procedure followed by the Army Trade Test Division in the construction of oral trade tests has been described in several books and articles[1]. We shall, therefore, include in the present paper only a brief summary of the statistical method used in the treatment of test questions. This will serve as a basis for our critical discussion.

For each trade question on which standardization data were collected, a graph like that shown in Figure I was constructed.

The question is scored from 4 to 0, for a correct answer, 0 for incorrect, and intermediate scores for partially correct answers[2]. In the graph the scores are read in the column to the left, while across the top are arranged the four trade classes—Novice, Apprentice, Journeyman, and Expert. The number of men of a given trade status making each score is entered in the table. Thus, 2 Novices scored 4; 8 Apprentices scored 4; 18 Novices scored 0; and so on. Twenty men of each class of ability are represented in the graph. The average score for each class is obtained and entered at the foot of the chart. These averages are plotted and a curve drawn to show graphically the relation between trade status and score. On the basis of an inspection of these graphs questions were retained or eliminated.

From those questions which were selected as showing sufficiently "good" graphs, the ones were retained which, in combination, would give the best discrimination between Novices and Apprenties, Apprentices and Journeymen, and Journeymen and Experts. The final test consisted of some ten to twenty questions, the combined score on which was then used in classifying the individual tests.

We shall attempt to examine critically both the treatment of individual trade questions and the method of combining and using these questions. The first—the manner of evaluating questions—

---

[1] The Personnel System of the United States Army. Washington: *Adjutant General's Department*, 1919. Chapter 20.
J. C. Chapman. Trade Tests. New York: Holt, 1921.
Additional references may be found in the last mentioned book.

[2] Partial scores were rarely used in actual practice. None are present in our illustration. Throughout the present paper we shall entirely ignore partial scores and shall speak of passing and failing as the only alternatives.

requires a statistical re-interpretation. The second, which has to do with the form of the test as a whole, is principally a matter of convenience and practical usefulness.

The method of evaluating the individual questions appear at once to be open to criticism regarding the statistical logic employed. The procedure makes use of the wrong regression line in considering the relation of ability and score. The inquiry made by the technique used is: What are the average scores made by tradesmen of the several grades of *ability*? The proper question would be: What is the average (or most probable) trade status of tradesmen making a particular score on a question?

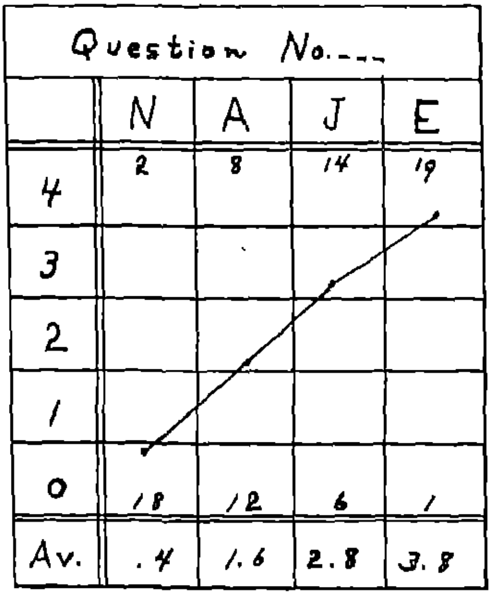


Figure 1.

On more careful examination, however, it will be found that while the procedure described does plot the wrong regression line, actually that regression relation is not used as a means of prediction. The regression *equation* that is, is not made use of either explicitly or implicitly.

In order to make clear the indirect use to which the regression of score on trade status is put, we shall first discuss the logic underlying the evaluation and combining of questions. It will then be noted wherein the Army method meets, and wherein it does not meet, the statistical requirements.

The problem which we have, if we ask how well we can actually predict trade status from score on the individual question may be formulated more definitely thus: What are the numerical odds that a man passing (or failing) a question is of one trade status rather than another?

A logical procedure for determining these odds follows:

(1) Let n^x, a^x, j^x, and e^x represent the numbers of Novices, Apprentices, Journeymen and Experts respectively who take a given question,x.

(2) Assume that $n^x = a^x = j^x = e^x$. (This is done in practice by using percentages, i. e., by referring results to a basis of 100 individuals in each trade class.)

(3) And let n^x_r, a^x_r, j^x_r, and e^x_r represent respectively the numbers of Novices, Apprentices, Journeymen, and Experts who give a "right" answer to the question x.

(4) Similarly n^x_w, a^x_w, j^x_w, and e^x_w represent the numbers who give a "wrong" answer to the question.

(5) Then clearly, the ratio of a^x_r to n^x_r expresses the odds that a man passing question x is an Apprentice rather than a Novice, assuming $n^x = a^x = j^x = e^x$.

(6) Corresponding statements hold for the ratios j^x_r to a^x_r and e^x_r to j^x_r, and for ratios using percentages of failures in place of successes.

Now these numerical expressions of the odds constitute precisely the sort of relations we desire save, of course, that we must in practice take into consideration the actual *inequality* of n^x, a^x, j^x, and e^x. The proper allowance for these inequalities can be made only on the basis of data which will indicate roughly, at least, what the proportion of Novices, Apprentices, Journeymen, and Experts are among men actually applying for employment in the given situation.

(7) If now, N, A, J, and E represent the actual numbers of

tradesmen applying,

(8) Then, $\frac{A}{N} \times \frac{a_r^x}{n_r^x}$ = the odds that a man who passes question x is an Apprentice rather than a Novice. That is to say, the formulation of (5) must be corrected by weighting the items in accordance with the actual number of Novices and Apprentices who are applicants. It is to be especially noted that for the purpose of comparing one question with another the weighting is wholly unnecessary. Since it remains constant, the relative merits of questions can be judged quite as well by using simply the ratios of the type, $\frac{a_r^x}{n_r^x}$.

The expression (8) stands, then, as the general answer to our question, by stating simply what the odds are, that a man passing a question is of one trade status rather than of another. Corresponding expressions may readily be written for other class differentiations and for situations in which the answer to the trade question is wrong instead of right.

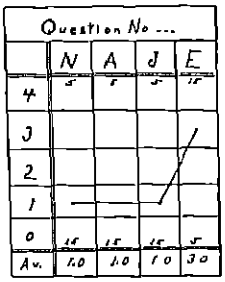
Let us now return to the Army procedure and examine it in the light of the preceding discussion. When the average score on a question was obtained for each trade class, what was really secured (as is evident from the nature of the computation) was the percentage of the class who passed, multiplied by a constant factor, 4. The regression line that was graphed, therefore, may be thought of simply as a curve showing the per cents of individuals passing at successive trade levels.

To the extent that the regression line or curve of per cents was of use in comparing the percentages of passes at adjacent trade status levels, it possessed a certain value in indicating the proper selection of questions; to the extent that it gave a false notion of the relative excellence of differentiation at the several levels, it was unsatisfactory. The sharpness of rise in the curve between two groups was considered the significant fact. The height of the point from which the rise takes place plays an important part which was not sufficiently considered in the Army procedure. Consider this point in greater detail. As indicated in our theoretical discussion, the critical expressions for determining differentiation are the ratios of the type, $\frac{x_{a_r}}{x_{n_r}}$. This ratio *can* be obtained directly from the curve plotted for each question in the old procedure. For example, (See Figure 1) if the curve rises from .4 (Novice average) to 1.6 (Apprentice average), the desired ratio is 4 (i.e., 1.6 divided by .4). In Army practice this ratio was not read. What *was* taken as a criterion (usually implicitly, through estimated steepness of the curve) is the *absolute* difference between successive points on the curve in the example just given, 1.2. Suppose, now, that the next point on the curve, the Journeyman average, was 2.8. The absolute difference and the steepness of the curve remain constant; but the ratio which was 4 at the first division is 1¾ at the second. Actually, that is, the question differentiates very much better between Apprentices and Novices than between Journeymen and Apprentices, although the Army method of analysis shows equally good differentiation.

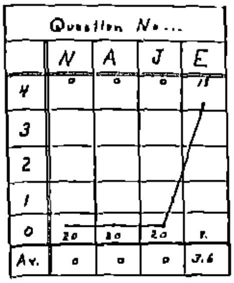
The shortcomings of the Army procedure can, perhaps, be made clear by reference to a few sample graphs of questions. These are shown in Figure 2. A man passing question B receives the same credit as one passing A, even though passing B is almost certainly diagnostic of Expert status whereas half the men passing A are not Experts, one sixth actually being Novices. Question C is considered a good Journeyman question; D is rejected. Note, however, that passing of D means that a man is practically certain to be either Journeyman or Expert while passing C gives a much lower probability. The same is true for failure on the question. E illustrates the sort of question discussed in the preceding paragraph. A man passing this question is four times as likely to be an Apprentice as he is to be a Novice, while he is only two and a half times as likely to be a Journeyman as an Apprentice. Nevertheless, this question is considered a Journeyman question simply because the curve rises more sharply at that level.

In concluding this discussion of the Army procedure a word s to be said regarding the satisfactory (on the whole) results btained. Two prime requirements for Army trade tests were speed of production and simplicity of administration. These requirements were met by the technique used. Any slight gains in ccuracy that could have been made through more refined statisical treatment, it may well be argued, would not have been worth the trouble. The fairly good results that were obtained by rough methods were in part due to the following considerations: (1) Questions were combined that did show good differentiation at the several trade levels. The differentiation was satisfactorily shown by the inexact method for the reason that ordinarily very good differentiation occurred at only one class division and this division point could be detected by merely noting the rise in the curve. Hence good results were secured though it seems reasonable to believe that the same quality of results might have been obtained with shorter tests and less waste of questions. (2) Since a fairly large number of questions was asked, the errors made in crediting separate questions were likely to be compensating.

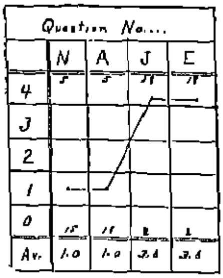
A.



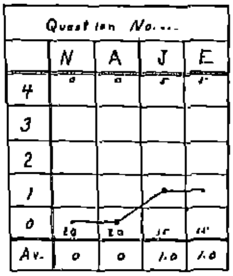
B.



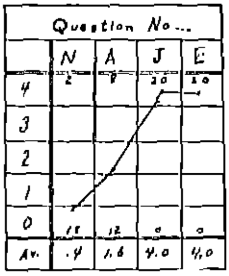
C.



D.



E.



The conclusion that the Army method of evaluating questions was at fault and that the practical choice of questions was to some extent adversely influenced by the statistical method utilized appears to be justified. The question at once arises whether a more satisfactory trade test method can be built, using the sort of statistical foundation we have outlined. The remainder of the paper will be devoted to some tentative suggestions for a type of trade test which does make use of the method presented for the evaluation of single questions, and which will at the same time have certain advantages as compared with the Army technique in the manner of the combining and using of questions.

The Army oral trade test consists of a fixed list of questions with norms established for the total list. A method of test construction which would make possible the treating of individual questions as standardized units, in place of having the entire test as the standardized unit, would have decided advantages in point of practical usefulness. Among the advantages are: (1) The test may be made more brief when desirable. (2) It may be varied to prevent coaching and to allow for unusual circumstances. (3) In testing the examinee it is unnecessary to give questions of all degrees of difficulty; that is, questions less difficult than those already passed need not be given, nor questions more difficult than those already failed. (4) The test may be continually developed by the incorporation of new questions, standardized in connection with the use of the existing questions. (5) Questions which prove undesirable for any reason may be dropped without necessitating any change in the remainder of the test.

In addition to these practical advantages listed, a system of trade test construction based upon the theoretical discussion in Part II, would have a more logical statistical foundation than have Army trade tests.

What, now, are the difficulties in the way of a new trade test method of the sort we have been pointing toward? Is it possible for us to lay down some definite suggestions as to the form of a trade test built on the single question basis and to outline a mode of procedure? The answer is that there exists one most formidable obstacle, namely, the absence of a sound method for combining the results from the single questions. Concretely, how shall we combine the ratios showing the odds in an example such as the following? Let us say the applicant is given five questions, three of which are passed and two of which are failed. We have the following odds from the several questions: Question 1 (passed) shows that his chances of being Apprentice rather than Novice are 9 to 1; question 2 (passed) shows that his chances of being Journeyman rather than Apprentice are 7 to 3; question 3 (failed) shows that his chances of being Expert rather than Journeyman are 2 to 8; question 4 (passed) shows that his chances of being Journeyman rather than Apprentice are 8 to 2; and question 5 (failed), shows that his chances of being Journeyman rather than Apprentice are 3 to 7. What, now, are the probabilities that this individual is a Journeyman, judging on the basis of all the questions?

The usual way to combine probabilities in answer to such a question is to multiply the ratios. Thus, in coin tossing, for example, the chances that one coin will come up heads is 1/2; the chances that two coins will come up heads at the same time is 1/2 x 1/2. Applying this sort of method to the above example, we would obtain the combined odds from the five questions by multiplying the ratios given. We would have, that is $\frac{9 \times 7 \times 2 \times 8 \times 3}{1 \times 3 \times 8 \times 2 \times 7} = \frac{9}{1}$. But this gives false results. This multiplying of probabilities is based upon the assumption of independence (lack of correlation) between the several events. However, the passing and failing of particular trade questions is clearly not uncorrelated with the passing and failing of other trade questions. It might appear at first that we could correct for the intercorrelations among questions by means of the partial correlation procedure. This, however, is impossible because of the inapplicability of the partial correlation method in the case of non-linear regressions. That the regressions of trade questions are not linear was clearly demonstrated by Army experience. The writer knows of no way by which this problem of combining probabilities from different trade questions can be handled mathematically.

We apparently, then, are forced into a position where we must either abandon the proposed sort of trade test or find some makeshift for combining questions. We shall briefly state one or two tentative suggestions in the latter direction. An empirical procedure for combining individual questions, even though it fall far short of what would be theoretically desirable, may nevertheless be sufficiently valuable to warrant the use of the proposed sort of trade test.

We shall mention two possible lines along which an empirical procedure might be developed. The essential requirement is that due provision be made to guard against the use of many closely intercorrelated questions in the examination of an individual. This may be accomplished either (1) by classifying appropriate questions into fixed groups placed at several levels of difficulty (analagous to Binet age groups), or (2) by having at each level a large number of questions which are sorted into classes of questions having relatively low correlations with questions in other classes.

The first of these suggested possibilities has some, but not all, of the advantages earlier mentioned as characteristic of a test which would flexibly combine single standardized questions. In this method we would have a compromise between the standardized single question and the standardized complete test. Here, the whole group of questions at the Apprentice, Journeyman, or Expert, level must be treated as a unit. This trade level form of test, while somewhat similar to the Binet age level method, has several significant differences. Among these are the following: (1) Questions are placed at the level where they differentiate best, not where a certain per cent of the individuals tested have passed. (2) Questions are weighted in proportion to their differentiating value; i. e., on the basis of the probability ratios earlier discussed. (3) Questions are weighted differently if passed and if failed. (These modifications would appear to be applicable to age-level tests like the Binet as well as to trade tests.)

The second suggestion of a procedure for trade test construction has certain advantages over and above those of the fixed trade level grouping just considered. It is essentially the same, save that greater flexibility is introduced in the use of questions. It carries very much further the idea of alternative tests which have been placed at some of the age-levels in revised Binet tests. In place of having a fixed set of questions at, say, the Journeyman level, we have a great many questions. These are classified according to the divisions into which trade knowledge in this trade has been analyzed. In examining a candidate, the questions which are asked at a given trade level are chosen from the several classes into which the questions at that level have been distributed. This method appears to be a feasible empirical device to prevent a candidate's being judged upon his answers to a non-representative and highly intercorrelated set of questions.

We shall not enter into a description of the detailed methods of weighting, scoring, and combining scores, which would be used in the suggested procedure. Suffice it to say that the actual probabilities calculated for each question can be transmuted into simple "weights" (consisting of numbers proportional to the logarithms of the ratios of odds) which may then, at each level, be readily added. The sum of the "weights", regardless of the number of questions asked, will give a rough indication of the odds that the candidate is at that level or below it, based upon all the questions asked at that level. The questions are, of course, in this way being used as though they were not intercorrelated. For adding the logarithms is equivalent to multiplying the ratio and this is the theoretically correct method of combining probabilities when the events are uncorrelated. The composite probability that we get is, therefore, only a very rough approximation.

But inasmuch as the questions have been classified into groups not closely correlated with questions in other groups, and inasmuch as the combined probabilities will be thrown off in about the same degree and in the same direction for all candidates, the procedure appears to be usable. We again emphasize, however, that the method is offered only as a tentative suggestion, a makeshift procedure which may have advantages over the rigid type of Army trade test.

IV.

We may briefly summarize as follows:

1. The statistical procedure used in building Army trade tests appears to be open to criticism on the grounds that the regression of test score on trade status was used in place of the reverse relationship. Closer consideration reveals, however, that the regression relations which were calculated were not used as such.

2. In order to make clear the indirect use of the figures calculated in evaluating each test question, we have outlined, without reference to Army procedure, a logical basis for evaluating questions in terms of the probability relationships of trade status to scores on single questions.

3. The Army method was then compared with our discussion of a method which would use mathematical odds for each question, and the conclusion was reached that to a certain extent the Army procedure was actually comparing such percentage ratio, but not in a logically correct manner. The method used worked tolerably well, but, as shown by a few illustrations, it also failed in part.

4. Finally, we have attempted to suggest lines which a different technique of trade test construction might follow, based upon the theoretical discussion contained in the paper. The principal advantages of the proposed sort of test are indicated as well as the special difficulties involved.

A CASE OF SECONDARY MENTAL DEFICIENCY WITH MUSICAL TALENT

By Blanche M. Minogue

Before medicine and psychology had advanced to their present degree of development, the obviously defective in intelligence were regarded as being possessed by some superior power. The occasional "idiot-savant" with his extraordinary abilities must have seemed convincing proof of that theory. Modern psychological methods, however, enable us to understand to some extent these heretofore inexplicable creatures. In recent years several cases have been reported in considerable detail. They are usually moderately high grade defectives who possess special abilities, which stand out in startling contrast to learning ability, judgment, etc., which are noticeably below the average.

Among "idiot-savants" are occasional musical geniuses who are almost phenomenal. Although musical genius in itself is apparently something apart from general intelligence, it involves many of the attributes of intelligence. For this reason a case of marked musical talent combined with low general ability is the more remarkable. Such a case is that of a male patient, at Letchworth Village at the present time.

X. Y., a twenty-three year old boy, with a mental age of about seven years, is a pianist of marked ability. He has been in the institution for several years, and in that time has played at practically all of the social functions of the institution.

The family history seems free from any intelligence defect. The paternal line is of normal, if not superior, mentality. The father's father was a broker, and one cousin, an engineer, has been a successful inventor. The father, born in Hungary, earns a moderate salary as a shipping clerk. The father's mother is a pianist of exceptional ability. His sister's daughter is also an unusually gifted pianist. The other members of the family, including the father himself, show musical ability in moderation.

The maternal line is interesting. The mother, also born in Hungary, is a woman of superior intelligence and refinement. The entire family appear to be of more than average mental ability. All have achieved success in business. One brother has two children who were pronounced superior after a psychological examination. This branch of the family shows considerable emotional instability, and does not give any indication of musical talent.

The patient's sister is a very bright, attractive young woman of about twenty-one years. She graduated from high school, winning recognition for her scholarshp. At present she is working and carrying college work in Extension School. She is a violinist, but does not show unusual talent.

The patient was born in 1900, in the United States. Everything connected with pre-natal life was favorable. Birth and infancy were normal. He was a bright, healthy child, walking at twenty months and talking at one year. As soon as he talked he learned little songs in English, German, French and Hungarian. He learned quickly and appeared to have intelligence, as well as musical ability, far above the average. Even in infancy, however, the child had periodic attacks of temper. His mother describes him as very amiable until some emotional experience such as surprise caused a sudden passionate outburst.

The boy developed normally until the age of three years. At that time he was severely ill with pneumonia, with complications which have since been diagnosed as meningitis. In the delirium of the disease he sang incessantly the songs he had learned. For many years after his recovery he showed great reluctance to sing. After this illness the child frequently had violent screaming spells and soon showed mental retardation. When seven years old he was taught at home for a year by a special teacher, and then entered grade 2A in the public schools. He made fair progress in reading, writing and arithmetic, but was teased so unmercifully by the other children that he had to be removed from school.

He received piano instruction for two years under excellent teachers but was very abusive to them. As he became older he often became voluble on the streets, and attracted crowds of people. He became the victim of unscrupulous persons in the neighborhood and so was committed to Letchworth Village.

Since admission to the institution X. has not been a discipline problem. He does light work around the cottage, and plays the piano much of the time. As he is not very robust he requires sleep during the day. He lacks many characteristics of the feebleminded. His muscular coordination is very good; he walks briskly, without the usual shuffling gait of the mental defective.

X. takes great pride in his musical accomplishments. He plays the most difficult music by sight and by ear. His ability to render classical and concert pieces, as well as the modern "jazz," is noteworthy. In a recent test before a woman of considerable musical training he played the *Anvil Chorus* from "Il Trovatore," *Lohengrin*, the Tannhauser *March*, *Tarantella*, *Carmen*, Chopin's *Funeral March* and many other classics. He read at sight the *Marche*

Grotesque by Sinding, Chaminade's *The Flatterer*, and *Heart's Melody* by Engelman. He then accompanied the phonograph in playing *Humoresque* and the *Blue Danube Waltz*, and had no difficulty in selecting the right key. He also read at sight the accompaniment for a vocal soloist. His emotional instability obviously has an unfortunate effect upon his playing, but he plays accurately and with feeling when his attention is focused upon his music rather than upon the admiration he is exciting. The music instructor at the institution considers his technique very fine for his limited period of study. His pitch discrimination is remarkable, and his tonal memory unusually good. His sense of rhythm seems to be more affected by his emotional state than are his other abilities, and consequently is more variable. The most important single factor in the boy's talent seems to be his memory. For time, for places, for events, for any composition ever learned, he has a memory almost phenomenal. That he has kinaesthetic memory also, is evidenced by his habit of playing without looking at the piano. His ability to imitate is considerably beyond his mental age. In the Knox cube test he made no failures.

Unfortunately, this patient is deteriorating mentally. On his admission to the institution in 1914, age fourteen years, his intelligence quotient was .62, at twenty years of age, his I. Q. was .54 (mental age eight years and six months) and a recent test gave him an I. Q. of .46 (mental age seven years and five months) by the Stanford Revision of the Binet-Simon. In this last test his basal year was four. At the five year level he failed the "patience" test, and in the sixth year passed all but the mutilated pictures. At the seventh year he failed to give differences and received only half credit for tying the bow. At eight years he failed the ball and field and the comprehension tests, and could not count backwards. At nine he knew the date, could make change and give rhymes. The subject failed all tests at the ten year level. The greatest deterioration appears to be in comprehension and judgment. At present he seems to fall at about the six year level in tests of these functions. In Burt's reasoning tests he made a score of 1.75, giving him a rating just below the seven year level. His score in the Indiana University Primer scale was 42, (the median for six years is 43.3). He was unable to go above the fifth year in the Porteus Maze tests.

Emotionally, this boy is excitable, egocentric, usually cheerful and obedient, but very sensitive and wholly incapable of a social existence. He does not mix with the other boys, and leads a life really apart from them, although one feels that he craves comradeship, but does not know how to get it.

There are two particularly interesting features of his talent. One is his failure to produce any original composition. This may be the result of defective intelligence. The other is his inability to learn to dance. He seems to have no comprehension of the relation of the musical rhythm to his own movements. This apparently is not dependent on intelligence, for many low grade defectives are excellent dancers.

The study of this case is by no means a complete one but it serves as one more bit of evidence that musical ability is a thing apart from "general intelligence." It is interesting that the disease which robbed the child of his heritage of intelligence and increased his emotional instability, left apparently intact this special talent. It is perhaps purely a matter of chance that he was left with one intellectual trait, his unusual memory, unimpaired. This fortunate combination of memory, sense of rythm, pitch discrimination, etc., makes a striking talent. Were any one of these lacking, or destroyed, the whole gift would go with it.

THE FRANZ DOT TAPPING TEST AS A MEASURE OF ATTENTION[1]

By Irene Case Sherman

In his "Handbook of Mental Examination Methods"[2] Dr. Franz describes various tests of attention, among them a test of the performance of accurate movements. Cross section paper (2 or 4 squares to the inch) is used; "The subject is instructed to tap in each square as rapidly as possible, going to and fro."[3] The experimenter notes the time for the taps along the particular lines. The test may be made with 300 squares, 20 squares to the line. Emphasis is placed neither upon accuracy nor upon speed, since the aim is to discover variations in attention.

In a personal communication Dr. Franz advised the use of 10 squares to the line with 30 lines for children. He states that the individual sets his own speed, but is expected to keep up the same speed throughout the test. With a speed not the maximum the rate can be kept up constantly with a variation of not more than 10%. He believes that the test shows ability to keep up progressive work, and also indicates "stick-to-it-ness." He also points out that the test shows the relative performance of the individual at different times, and does not give a comparison of different individuals.

In an effort to find an accurate means of measuring fluctuations of attention in children, this test, with a number of so-called attention tests, was given to 172 unselected cases[4]. Cross section paper four squares to the inch was used, with ten squares to a row in a series of thirty rows. The instructions to the child were, "Put a dot in each of these squares, beginning here and going across to here (examiner indicated the distance of the ten squares), then coming back on the next row, and so on down to this line (indicating the base line which marked off the thirty rows). Do this as fast as you can, but be sure to put a dot in each square, and don't miss any squares." The time for completing each row of squares was taken.

The results indicated such a wide variability of performance in children that it was felt to be necessary to establish a more accurate norm than the 10% variability suggested by Dr. Franz, provided the results would lend themselves to such a standardization.

---

[1] Dr. Franz does not name this test, but the writer feels "Dot Tapping" is most descriptive of it.

[2] Franz, Shepherd Ivory, *Handbook of Mental Examination Methods*. The MacMillan Company, 1920.

[3] Ibid. P. 80.

[4] Institute for Juvenile Research, Chicago, Illinois.

Each child was graded by the Stanford Revision of the Binet test, so that a mental age and intelligence quotient was available in each case. In each case the arithmetic mean, the range and the average deviation of time were taken, as well as the coefficient of variability. In order to discover the influence of age, mental age and intelligence rating upon performance in the test, the results are arranged according to these three factors, and are presented in tables I, II and III.

Table I indicates that there is a general decrease in average time, average range and average deviation with increase in chronological age. This decrease, however is not sufficiently regular to make the chronological age a basis of standardization of performance, and is no doubt due to the increase in speed due to better motor co-ordination which develops in advance in age. The coefficient of variability showed an even greater variability in relation to age, so that it indicates no relation between variability and chronological age.

TABLE I

RESULTS ACCORDING TO CHRONOLOGICAL AGE

Age	Average Time	Average Range	Average Deviation	Coefficient of Varinbility	No. of Cases
7–0	11.27	13.9	2.26	19	9
8–0	12.78	8.9	1.65	14	7
9–0	7.85	7.2	1.37	18	15
10–0	6.84	6.6	1.17	16	20
11–0	6.72	6.1	.98	13.5	19
12–0	5.62	4.9	.92	16	21
13–0	5.65	5.1	.99	17	23
14–0	3.96	3.4	.5	12	26
15–0	5.16	3.6	.70	11.68	15
16 +	5.11	3.7	.77	15	17
TOTAL					172

Table II shows the same tendency for average time to decrease with increasing mental age, the average range and average deviation also show this decrease, but not to such a marked extent as average time does. While this decrease in average time is quite definite, yet the differences between any two mental ages are too small to make the average time a basis of scoring in terms of mental age. The coefficient of variability, as in the case of chronological *age, shows great variation, and no relation to mental age.*

Table III indicates that none of the measures showed any relation to intelligence quotient.

Table IV and graph 1 give the distribution of the coefficients of variation for the entire number of cases. The median coefficient is 14.75 and the inter-quartile ranges from 11 to 18 inclusive. That is, fifty per cent of our cases have a coefficient of variation between 11 and 18. A score falling between these numbers, then, indicates a normal performance in this test. The upper quartile *ranges from one through ten, and any score between these numbers* would be considered an excellent performance in this test. A score of nineteen or above is considered to show a fluctuation of attention greater than that shown by fifty percent of our cases.

Graphs 2, 3 and 4 illustrate the increased variability of performance of individuals with a coefficient of variability above 18. The time is more variable in subjects with a high coefficient, and they

TABLE II

RESULTS ACCORDING TO MENTAL AGE

Age	Average Time	Average Range	Average Deviation	Coefficient of Variability	No. of Cases
5–0	9.17	6.0	1.16	13	1
6–0	13.52	10.3	1.98	15	11
7–0	10.45	13.7	3.17	23	16
8–0	6.97	5.2	1.01	14	26
9–0	6.31	5.3	.97	15	28
10–0	5.42	5.5	.89	16	32
11–0	4.56	3.0	.63	14	14
12–0	4.85	3.3	.67	14	15
13–0	4.09	2.6	.43	11	14
14–0	4.03	2.8	.43	10	7
15–0	4.95	4.1	.89	17	4
16+	3.55	2.0	.34	10	4
TOTAL					172

TABLE III

RESULTS ACCORDING TO INTELLIGENCE QUOTIENT

I.Q.	Average Time	Average Range	Average Deviation	Coefficient of Variability	No. of Cases
50	7.0	5.9	1.5	15	9
60	7.1	6.5	1.3	17	23
70	6.6	5.4	1.0	16	41
80	8.4	6.8	1.2	14	43
90	6.3	7.4	1.2	16	26
100	5.0	3.5	.6	11	20
110+	4.9	4.0	.7	12	10
TOTAL					172

TABLE IV

DISTRIBUTION OF COEFFICIENTS OF VARIABILITY—172 CASES

	0- 1	1
	1- 2	1
	2- 3	2
	3- 4	1
	4- 5	2
	5- 6	3
	6- 7	5
	7- 8	8
	8- 9	5
	9-10	12
25 percentile	10-11	16
	11-12	12
	12-13	12
Median	13-14	13
	14-15	8
	15-16	8
	16-17	12
75 percentile	17-18	13
	18-19	7
	19-20	9
	20-21	6
	21-22	3
	22-23	0
	23-24	3
	24-25	3
	25-26	3
	26-27	0
	27-28	1
	28-29	0
	29-30	2
	30-31	0
	31-32	0
	32-33	0
	33-34	0
	34-35	0
	35-36	0
	36-37	1
	37-38	0
	38-39	0
	39-40	0
	40-41	1
		172

equire on the whole a greater length of time to complete the est than subjects with average coefficients. This difference is nore marked in the case of the average deviation, where the neasures are greatly dispersed for those coefficients above 19, vhile those below 18 show a central tendency around a low iverage deviation. The mode for those making an excellent score (3–10) is .1 while the mode for the average group is somevhat higher, namely .7. The range of time taken on the test is also higher for those with coefficients above 18, being more widely distributed than the measures for the lower coefficients, which as in the case of average deviation, show a marked central tendency around a low score. These graphs clearly show that a high coefficient of variability is accompanied by increased total time of performance, increased range of time from line to line, and increased deviation from line to line. Since it has been shown above that the coefficient of variability is independent of age, mental age and intelligence quotient, it may be assumed that the variability of the coefficient is due to some other factor, presumably fluctuations of the attention.

Graph 5 illustrates the typical kinds of performance on this test. The first curve has a coefficient of variability of 4, which is superior performance. The average time is five seconds, with long periods of no variation from this average time; there are no great deviations from the average from line to line, and the total range of time is only two seconds. The second curve is that of an average performance, with a coefficient of variability of 14. The average time is also five seconds, but there are no long periods of unvaried performance, although on the whole the deviations are slight. The total range of time, however, is five seconds. Curve three is a curve typical of a high coefficient of variability, in this case 20. The average time here is 13 seconds, with no periods of invariable performance. There are constant deviations from the average from line to line, with a wide total range of time, namely, 15 seconds.

In the giving of the tests it was often noticed by the examiner that children frequently varied the size of the dot, began making circles, heavy dots, two dots in one square, etc. This was often interpreted as due to low chronological or mental age, or to low general intelligence. Since the results showed no relation between these factors and the coefficient of variability, a study of the dot patterns was made. This revealed that the variability of the size of the dots had little relation to the score. Subjects who changed from a large dot to a small dot, or vice versa, did not necessarily make a poor score. In fact, many changes in the size of the dots occurred with no change in time from line to line. On the other hand, many subjects who made very poor scores showed a marked regularity in the size of dot. In only one case was there a definite relation between size of dot and increased time. This case, scoring 18, had a marked increase in time in two lines, with a change in dots in these lines, making circles instead of dots in these two instances. On the other hand, of six cases showing lines with marked increase of ime with increase in size of some of the dots in this line, all of hese cases also had other lines with increased rate of time and no hange in size of dots or other variations. While it is possible hat a change in the size of the dot may mean a shift of attention, t is almost inpossible to say definitely that this is true, because f the large number of cases in which there was no evident reation between time and size of dot. Changes of time are just as ikely to occur where there are no changes in size of dot as where here are such changes, and on the whole these changes in size f dot are considered as insignificant compared with other unnown factors which are producing changes in attention. In fact, t is believed that these changes in size of dot are but one of the xternal manifestations of the variability of attention, relatively nimportant, since they seem to influence the total score very ittle.

The records of the cases were analyzed in order to discover, if ossible, any factors in the subject which might cause, or be corelated with unstable attention. Because of the subjective nature f the analysis of the individual cases, with the marked lack of orrespondence of descriptive terms used in the various records n presenting the history and mental analysis, a limited number f terms were arbitrarily selected as possibly having some reation to attention. In cases designated as negative none of these descriptive terms occurred.

In Table VI the individuals are classified under the descriptive erm which most nearly described them; the percentage distriution is given according to coefficients of variability. The table hows that the smallest percentage of individuals with a negative history occurs in the group with scores of 19 or above. This same roup also showed a larger percentage of "irritability," "distractbility," "incoherence," and marked "slowness of reaction" than did the other two groups. Subjects described merely as nervous nd restless were found not to show poor attention as measured by his test. It was also striking that individuals described as being oor in attention by various examiners or other observers in the ome or school were not characterized by making an unusually high core, i. e., 5% of these subjects made coefficients from 3 to 10, and ikewise 5% made coefficients of 19 or above. Similarly, subjects with organic brain disease or psychoses were not characterized by high score. Rather, the larger percentage of these were found to make an excellent score, namely 3 to 10, and were able to sustain ttention over a short period of time, such as is demanded in this est.

TABLE VI

PERCENTAGE DISTRIBUTION OF TYPE OF CASE, ACCORDING TO COEFFICIENT OF VARIABILITY

Coefficient of Variability	Negative	Poor Attention	Nervous and Restless	Unstable	Irritable	Hyperactive, Distractible and Flighty
3-10	42.1%	5.3	18.4	7.9	2.6	2.60
11-18	49.5%	3.2	9.5	1.0	9.5	3.21
19 +	28.2%	5.0	12.8	7.7	12.8	7.70

Coefficient of Variability	Rambling and Incoherent	Slow	Apathetic	Psychotic	No record	Total
3-10	2.6	2.6	5.3	7.9	2.6	99.9
11-18	*0*	*0*	*12.6*	*2.1*	*8.4*	*100.0*
19 +	5.1	7.7	5.1	2.6	5.1	99.9

CONCLUSION

While average time, range and deviation decrease with increasing chronological and mental age, this decrease is not sufficient to make chronological or mental age a basis of standardization, since the increase is probably due purely to increasing motor coordination which develops with increasing maturity. The better performance with increasing age, then, is not due to better controlled attention, since the variability of performance as measured by the coefficient of variability shows no relation to age.

There is no relation between performance in this test and intelligence quotient. Variability on the test is due to some factor other than maturity or general intelligence, presumably fluctuation of attention.

The inter-quartile range of coefficients of variability for this group of 172 unselected cases was 11–18. A score within this range is considered average performance.

A high coefficient of variability is accompanied by increasing total time of performance, increasing range of time, and increasing deviation.

Among cases showing a high degree of variability (score of 19+) there was a greater percentage found to be "irritable," "distractible," "incoherent" and "slow in reaction." Subjects described merely as "nervous and restless" were found not to show poor attention as measured by this test. Performance on the test showed that children described by observers in the home or school as inattentive did not necessarily make poor scores on this test. Observed inattention in these cases is evidently not a real lack of ability to attend, but is expressive of some other factors.

Similar results were reported in a recent article[5] on post-encephalitic children, where it was found that, although these cases were described by observers as inattentive, by tests they proved capable of sustained attention. The writers concluded that in their cases the apparent inattentiveness was the result of emotional disturbances[6].

---

[5] Sherman and Beverly. *The Factor of Deterioration in Children Showing Behavior Difficulties Following Epidemic Encephalitis.* Arch. of Neur. & Psych., 1923, Vol. 10, pp. 329-340.

[6] The writer is indebted to Myrtle Raymaker Worthington for assistance in preparation of the data.

N ADVANCED COURSE IN EXPERIMENTAL PSYCHOLOGY APPLIED TO ADVERTISING[*]

By A. J. SNOW
Northwestern University

It is obvious that the teaching of advertising without an ex-rimental laboratory is just as futile as teaching civil engineering thout a field course in surveying. However, there is a vital ference between the two: in the latter case an engineering ident can check up the correctness of his surveying with the l of mathematical computation; while in the case of psychology applied to advertising the student, at the present, cannot rify his theories except by reference to a purchasing public to iich he has usually no access. Nevertheless it is perhaps pos-le to bring this contact about—between the student of advertis- g and the purchasing public—by a cooperative plan on the part the advertiser and the university.

Obviously a student who would profit by such a cooperative rangement cannot be a beginner in the field. It is vital that should have some knowledge of the fundamentals of advertising fore he is admitted to such a course. It has been our policy to ke the following prerequisites for the course in experimental vertising: one semester of general elementary psychology, one marketing, and another semester in psychology as applied to ad-rtising. We are hoping that, in the last-named course, the student ll become familiar with the application of psychological principles the general field of advertising. To accomplish this end, the ident is asked not only to read published material and listen any new ideas brought out in the lectures, but also, as part his home work, to criticise advertisements, select samples advertising illustrating good and bad applications of psychologi-l principles, or lack of application, discover modern tendencies d solve any problems which were formulated by the instructor the purpose of applying psychological principles to advertising. me of these problems were of the following nature: one such oblem was as follows—Explain why advertising is limited to use of only two of the four possible ways of producing tention. Illustrate. For the second example take the case of Illinois farmer who wishes to advertise in the weekly news-

---

Read before the annual meeting of The National Association of Teachers Advertising meeting with The Associated Advertising Clubs of the World, antic City, June 6, 1923.

papers in order to market by express some new potatoes direct to the consumer. Three good papers are available, each covering different territory. A full page of space in any of them would cost fifty dollars, smaller space in proportion. How can the farmer most profitably spend fifty dollars in advertising in a given week? Why? Here is still another problem—An advertising man lost his job, the reason being that he arranged for an advertisement to appear every other month in a monthly magazine. The excuse he gave was that the appropriation was limited to the cost of only six issues. Why did this advertising man get the worst of it?

When a student has successfully completed this comparatively elementary course in advertising he can, in his senior year, or while pursuing graduate studies, elect the advanced course in experimental psychology applied to advertising. He can pursue this course either one or two semesters, gaining academic credit proportionately to the work done, but not exceeding five credit hours during any one semester. The students in this course meet twice a week in one hour sessions for discussion and twice a week in three hour sessions for laboratory practice.

The course itself is divided into two parts—individual problems and class problems. The first part—individual problems—is divided further into two parts, the division being based on the individual ability of the students in the group. The more promising students are assigned problems of minor research, individually or in groups, while the less capable students are assigned to perform standard laboratory experiments. For example, the following two are some of the problems of minor research assigned to the more capable students:—the first of these is an experimental study of the relative persuasiveness of types of advertising. Fifty types of advertising appeals were tried upon twelve hundred day and night school students. The second problem was a study of the legibility of printing types as classified in groups with relation to their feeling tone. The less capable students were instructed to duplicate some of the well known experiments in the field of the psychology of advertising devised by such men as Scott, Gale, Hollingworth, and Strong.

The second part of the course, required of all students, aims to supply the need of an actual contact with the buying public, and also attempts to bring before the students *practical* problems of advertising. In this, the *applied* part of the course, the students were concerned with the actual field problems in advertising. While it is, *strictly* speaking, impossible to separate the actual business part of advertising from the application of psychological inciples to advertising, *practically* at least, that can be done in measure. The reason for such an arbitrary division, into the isiness and the psychological aspects of advertising, is that a urse in psychology cannot teach the student how to run the isiness as a whole but only how to apply psychological prin- ples to certain aspects of it.

The success of this applied part of the course depends upon the ility of the instructor in charge, or the university as his agent, obtain the good will and cooperation of two important groups: st, small scale advertisers; second, a group of representative ofessional advertising men. The vicinity of the university to large industrial city, while not absolutely essential, lends itself lendidly to this purpose.

Again in this *applied* part of the course the class helps to termine the nature and function of the advertising which is bmitted for its judgment by the small scale advertiser. In that ay the class has the opportunity to verify its experiments and e validity of its claims for the application of psychological tegories to the advertising of merchandise, just as much, as the gineering student is able to verify his surveying data mathemati- lly. In order not to go far wrong in the decisions of the class, id in this way waste time and money for its clients, the problems fore the class and their decisions are discussed by the instructor om time to time with the above-mentioned advisory board com- sed of professional advertising men of high standing, chosen such a way that they as a group represent the various types advertising agencies now in existence.

It is important to understand that the students in experimental lvertising are not expected to represent artistically their ideas the making of the usual cut and copy. That is done for them id upon their bidding by professional artists and copy writers. is not the function of this class, or any other in the field of sychology as applied to advertising, to teach the technique of rrying out ideas in the form of cut and copy. The function ust always lie with the schools of art and expression, and only ith those students who possess such a capacity.

For the sake of clearness it may be expedient to state some of ie field problems which made up the task for the class:

First. A local storage and transfer company asked the class consider its advertising problem, especially during the months f April and May, September and October.

Second. A Chicago advertising agency requested the class to suggest the method and the nature of the copy to be used in order to put before the public a new type of baby carriage.

Third. A large manufacturer of paint asked the class, through his Chicago advertising agent, to prepare for his use a pamphlet of information about the nature and utility of his merchandise, to be distributed by mail, coupon attached.

Fourth. A manufacturer of garters requested the class to judge a secondary slogan with reference to its effect upon the potential buyer, and if necessary to construct a new slogan more suitable for its purpose.

This is by no means the only method that can be used to teach that sort of course, but it is a method, we believe, that has had a certain degree of success. The primary purpose of this paper, however, is to stimulate a vigorous discussion of the subject it presents.

[EASURING THE COMPREHENSION OF ADVERTISEMENTS

By A. T. POFFENBERGER
COLUMBIA UNIVERSITY

One of the most fundamental rules of a writer or a speaker is iat he shall "come down to" or not go "over the heads of his udience." The speaker may get an immediate reaction from his udience that will tell him whether or not he is obeying this rule. he writer, and especially the advertising copy writer, may not lways be so fortunate. His returns from any one piece or series f pieces of copy may be extremely slow and indirect. Where every ord costs many dollars, it is extremely important that each shall irry a message that will pay for itself and also bring a profit. o be sure, pictures will do much to carry the message, and may t the same time hide the unintelligiblity of the copy. But even iotion pictures require some easily comprehended copy to make iem acceptable.

Advertising, particularly national advertising, is mass selling nd to be effective must appeal to the bulk of the public. It is not asy to know the ability of this vast public to comprehend the rinted message. In fact, no one seems to have even guessed at ie level of intelligence, so called, of the white population of the Inited States until the results of the Army intelligence tests were eported. The percentage of illiteracy has received the most at-ention, altho the much larger percentage of low degree of literacy i quite as important. We need not here concern ourselves with ie heated discussions as to whether the average adult has an itelligence of a thirteen, fourteen or sixteen year-old person. Most f these discussions center around the question whether printed ymbols are adequate material for measuring native intelligence. Ve are here concerned with the capacity of the population to get he meaning of printed symbols, and with the degree to which ie advertising copy-writer keeps within the limits set by the capcity of the average person.

Now the average adult can read, understand and deal with irinted symbols about as well as the 14 year old child, or one vho has completed the elementary schools grades. It is to an udience of no higher type than this, about one half of the counry's population, that the advertiser of generally used commodities must appeal. That is, this degree of capacity sets the limit beyond which the copy-writer should not go.

1. To what degree does the copy-writer adjust his vocabulary to his audience?

With our little knowledge of the make-up of the advertiser's audience, complicated by the shifting of maximum buying power from the professional classes to the laboring classes, it is not safe to cling to an untested opinion on this matter. Moreover, it is a question which lends itself readily to experimental measurement. It one wants to know whether his copy is intelligible to his audience let him try it on a sampling of that audience. The following studies are intended to show whether or not such tests are necessary and at the same time to show how they may be made.

All the words, excluding the names of articles, were gathered from 190 advertisements appearing in one issue of the "Saturday Evening Post," "The Ladies' Home Journal" and the "Woman's Home Companion." There were in all 6378 words and when all repetitions of various sorts were eliminated, there remained 3796 different words. This list of words was checked against the words appearing in the "Teachers Word Book," prepared by E. L. Thorndike. This is an alphabetical list of the 10,000 words occurring most commonly in a count of about 625,000 words from literature for children and forty other sources. Of our list collected from the advertisements, only 4 percent failed to appear in the "Teachers Word Book." The copy-writer therefore, seems to keep fairly well within these limits.

Fifteen of the words appearing in both lists were chosen at random and presented to 37 persons, none of whom had gone through High School, who were asked to give their meaning. The percent of errors for the different words including no answers and wrong answers, ranged all the way from 2.7 percent for the word "reputation" to 73 percent for the word "zest." The average percent of error was 12 percent. Six words appearing in the advertisement word list but not in the "Teachers Word Book" list were now tried on the same 37 people. The percent of errors ranged from 0 for "parasite" to 60 percent for "slush." Although 37 people of elementary school grade do not form an adequate sample of any one class of the population, these figures suggest that even such a word list as that in the "Teachers Word Book" would not be an adequate guide for the copy-writer to follow if he is to be certain of understanding on the part of his audience. The percent of error was just as great for words appearing in the list as for those which were not in the list.

The following experiment with an actual advertisement showed e effect of the poor adjustment of an advertisement's vocabulary its audience. An advertisement of the poster type for Premium am contained the following statement: "An Epicure's Way of aking Ham." The whole appeal lay in the meaning and connotaon of the word "epicure" and must have been intended to produce favorable reaction in the lower half as well as the upper half of e population.

Another advertisement was prepared, duplicating this one in ery respect except that the word "Igorot" was substituted for e word "epicure." These two advertisements were shown to ch of 53 housewives with the requests that they state which ay of baking ham seemed to them the more desirable. Twenty ne chose the "epicure" advertisement while 24 chose the "Igorot" lvertisement, indicating that they preferred to bake their ham e savage way. Since it appeared quite likely that some of the who chose correctly may have been only guessing, they were sked to give their idea of an eqicure. Seventeen of them showed at they had sufficient knowledge to make a correct choice, hile the other 12 acknowledged that they knew the meaning of ither word and had no particular reason for their choice. They d chosen one merely to comply with the investigator's request. ixty eight percent of this sample group had failed to get the eaning which was intended. Although the illustration carried e atmosphere of a very pleasing food and made the advertiseent a good one, another more readily understood word would rtainly have added much to its effectiveness.

2 To what degree does the copy-writer adjust his ideas to his udience?

It is quite evident that a person may understand every word a sentence and yet not get the meaning conveyed by the sentence s a whole, may not get the idea in the mind of the writer. To etermine the degree to which the ideas presented in an advertiseent are grasped by the public a series of tests was conducted. hese tests took the form of the Thorndike Reading Tests which re designed to measure the degree to which one understands what e reads. A passage is presented to be read along with a series f questions to be answered from the material read. In our tests n advertisement in its complete form together with a series of uestions was used, or the copy of the advertisement was preented in typewritten form together with a series of questions earing upon it. Nineteen sets of material representing different dvertisements were tested on more than 1,000 people. The adertisements ranged in character from the simple Subway Car ard to the elaborate automobile advertisements. As it is impossible to present the data for all the advertisements in sufficiently brief space, samples will be chosen to illustrate the method and the nature of the results.

One hundred and seven New York Subway passengers were asked to read the advertising card for the .... Emulsion which contained the following copy:

"Milk is an emulsion. The .... Emulsion is emulsified 550 times finer than milk and its fat content is eight times richer than the best milk. That is why physicians and druggists all over America, when asked to name the finest emulsion, answer The .. .. Emulsion. Tastes good—No Cod Liver Oil-Protects and Builds Health."

The following questions were asked and the answers recorded by the investigator:

1. What is an emulsion?

2. What does "emulsified 550 times finer than milk" mean to you?

3. What is it that makes milk an emulsion?

4. Why does the ... . Emulsion protect and build health?

The answers to questions 1 and 3 demonstrated that none knew the essential characteristics of an emulsion, altho 46 percent knew that an emulsion was fatty. To others it was merely a liquid and to still others it was merely white. Answers to question two showed that none grasped the significance of the statement. About 30 percent did say that it meant that the .... Emulsion is 550 times better than milk. Although this belief is misleading it is not damaging to the product. Sixty-four percent believed that this emulsion would protect and build health, because "it is better than milk" (24 percent) or because "it is recommended by physicians and druggists" (40 percent). Although this advertisement may be considered effective, there is a great question whether the copy-writer really "put his ideas over," and obtained the maximum effect possible.

Forty seven women, all of whom looked like housewives, were asked to read the following car card, which is of the simple publicity sort:

"................ Golden Syrup. Made from Cane Sugar. For Table-For Cooking."

The questions asked were:

1. Why should you buy a can of .... Syrup?

2. Why does cane sugar make good syrup?

3. What other kind of sugar could be used to make syrup?

To question one, eighty percent said they would buy it because it was made from cane sugar. This was, of course, the only selling gument presented in the advertisement. But in answer to the cond question *it was discovered that no one knew why cane sugar uld make good syrup. The answers to question 3 showed that* percent knew of no other kind of sugar from which syrup ight be made. Further questioning of those who said they would y this product showed that it was the prestige of the trade me that was effective, and that the only actual reason given for ying this product had not done its work. .

Following is the copy taken from an automobile advertisement d presented to 55 college men with the request that they read and answer one question about it.

"Humanity may be a million years old in point of time, but is as young as this morning's sun in its pursuit of the ideal. fter two thousand years of disappointment and disillusion, the ernal verities and the eternal values still prevail . . . . Even ough it be surrounded and seemingly obscured by sham and etense, nothing in this world is discovered so surely as solid erit . . . . . This is the truth that embodies all truth; this is e *truth that makes men free.*"

Thirty percent of the group could not answer the question which as: What is the truth that embodies all truth? As they did t discover that "nothing in this world is discovered so surely solid merit," they could see no connection between this para- aph and an automobile advertisement. One person who had d experience in copy-writing agreed that "it might be part of poor advertisement for a good automobile." It seemed useless try this paragraph on less intelligent groups of people, who the same time might represent better automobile prospects an the group used in our experiment.

This paragraph was difficult or impossible to understand be- use of its abstract ideas and the complicated character of its nstruction. Another case studied, that of a doube-page adver- sement for a new safety razor owed its difficulty to the wealth technical details involving a greater knowledge of physics than possessed by the majority of people, even of college students *ho have had the advantage of a course in physics. The actual* dvertisement was used in this experiment and its content is o long to reproduce in this connection. It is described in some tail in another article by the author[1]. All that need be mentioned ere is that of the 55 college students who were asked to answer a ries of 7 questions about this advertisement, with the privilege f consulting it as often as desired, practically none was able to grasp the technical details on which the success of the product depended.

---

[1] A. T. Poffenberger, Journal of Applied Psychology, 1923.

The study of the nineteen cases, of which those given are samples, shows that advertising copy both in the vocabulary used and in the nature of the ideas presented frequently "goes over the heads of its audience." The trouble is frequently hidden in one or other of two ways or both. First, the excellence of the illustration may overcome the shortcomings of the copy or make the advertisement successful in spite of the copy. That copy can convey the most abstruse ideas so that they shall be easily understood, the reading of the masterpieces of literature will show. Second, people may be convinced of the value of an article even when they do not understand what is said about it. But since we do not know the extent to which this sort of reaction prevails in human behavior it does not seem safe to rely on it for successful advertising. It is a matter, however, which the student of Applied Psychology might well investigate.

The unfailing remedy for suspected unintelligible copy or copy that is difficult to understand is, as suggested in the beginning of this paper, a very simple one. It is, "Try it on the audience." Take a sampling of that portion of the public to whom the copy is to appeal and by a series of questions, carefully prepared beforehand, *find out just what is the reaction to the copy.* The Applied Psychologist who is interested in measuring and predicting human behavior in cases of this sort relies upon no general principles to guide him, but resorts to the method of sampling just as the chemical engineer or the physical engineer tests his materials by the method of sampling. And just as in these latter cases, skill is required in obtaining the samples, in measuring them, and in drawing conclusions from the measurements.

The advice of Robert Louis Stevenson to "Write so that the fool can understand you, and the wise man will understand you also," is, indeed, good, provided that one takes pains to discover what the fool can understand.

NEWS AND NOTES

The increased appropriation voted by the Eighty-fifth General ssembly of Ohio to the Bureau of Special Education and Psycho- lucational Clinic of Miami University, has enabled the Bureau increase its staff, to enlarge its facilities for training teachers d supervisors for special classes for subnormals, delinquents and eech defectives, and psychological examiners, and to open a anch station in Cleveland. In addition to the director, J. E. W. allin, the staff includes the following: in the Dayton center, Mrs. ssie Wilson, M. A., University of Michigan, and Miss Janet lcomb, M. A., Columbia University; Psychological examiners d instructors, Miss Mildred Grimes, B. A., Western Reserve and aduate student at Columbia, instructor in industrial arts, critic acher and Miss Eva Nutting, student in the Bureau of Special lucation, Miami University, critic teacher; and in the Cleveland anch, Bertha M. Luckey, Ph. D., University of Nebraska, and eva Witherstay, critic teacher.

Students are admitted from all sections of the country without her charges than a nominal registration fee.

SPECIAL EDUCATION, OHIO UNIVERSITY, ATHENS, OHIO

The work in Special Education was established in the College Education of Ohio University, Athens, Ohio, in 1920 in con- ction with the Department of Psychology and Paidology. In the ll of 1922 this work in Special Education was made a separate rt of the Training School in order to take care of the Special asses in the City Schools of Athens as well as to meet the calls ming from the students desiring practice in schools for excep- nal children. Miss M. La Vinia Warner, graduate of Columbia niversity, New York City, is directing the new department.

Special facilities were already afforded by Ohio University r the preparation of teachers for this line of work because of e opportunity for the selection of subjects offered by the llege of Education and also the College of Liberal Arts to hich are now added the Special subjects further needed and the pecial School for exceptional children as a training school. The ed for special education in the public schools of Athens and thens County affords a wonderful opportunity for the study of l types of exceptional children and the social conditions give udents a good opportunity for the study of and practical ex- erience in Social Work.

A group of fourteen students registered for the work in Special ducation in the 1922-23 term and thirty-seven students came in r the Summer Session of 1923. The demand for the work is growing rapidly throughout the State of Ohio and more calls than can be filled are coming in for teachers trained in this field.

W. V. Bingham has been elected Editor and L. L. Thurstone Associate Editor of the Journal of Personnel Research. C. S. Yoakum continues as Managing Editor of the Journal which is now entering on its second volume. The other members of the editorial board are Wesley C. Mitchell, Alice Hamilton, Frankwood E. Williams, R. W. Husband, Matthew Woll, Leonard Outhwaite, Joseph K. Willets, Lewis M. Terman, Alfred D. Flinn and Mary Van Kleek.

This Journal is devoted to the scientific study of personnel. It publishes original researches in the applied sciences that contribute to our knowledge and our effective direction of people at work. It is the official organ of the Personnel Research Federation whose purpose is the furtherance of research activities pertaining to personnel in industry, commerce, education and government wherever such researches are conducted in the spirit and with the methods of science. It brings together workers in the biological and medical sciences, psychologists, psychiatrists, engineers and economists who are making significant contributions to more exact knowledge of human beings at work and methods of making effective and satisfying adjustments between men and their work.

BRITISH ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

TORONTO MEETING, AUGUST 6TH TO 13TH, 1924.

SECTION J (PSYCHOLOGY)

Dear Sir (or Madam),

A meeting of the British Association for the Advancement of Science will be held at Toronto, from August 6th to 13th, 1924.

The association has recently added to its various Sections a Section of Psychology. The Organising Committee of the Psychological Section believes that the meeting will afford an unique opportunity for intercourse between psychologists of Canada, the United States, and Great Britain; and accordingly issues an invitation to any psychologist in the United States or Canada, who may desire to do so, to join the Association as a member for the Meeting, and to contribute a paper or attend the discussions.

Professor McDougall, F.R.S., of Harvard University, has accepted appointment as President of the Section; and it is hoped that he will deliver the presidential address.. Other papers have been promised by Dr. Morton Prince, Dr. Brett, Dr. Bott, Dr. Bridges, Dr. Sandford, Dr. Tait, by Dr. Myers, Dr. Burt, Dr. Drever, Mr. Flügel, Dr. Miles, Professor Pear, and other leading American and British psychologists.

Joint Meetings have been arranged with the following Sections:—

Section H (Anthropology) to discuss "Racial Mental Differences."

Section I (Physiology) to discuss "Application of Physiology and Psychology to Industrial Efficiency."

Section L (Education) to discuss "Tests for Scholarships and Promotion."

Papers should not require more than 20-30 minutes to read; and intending contributors should send a notice of their papers before the end of March to the Recorder of the Psychological Section (British Association), The Psychological Laboratory, The University, Leeds, England. Further particulars as to membership for the meeting can be had from the Secretary, British Association, Burlington House, Piccadilly, London, W.1., England; and notices of the general arrangement will be issued in due course by him and by the Hon. Local Secretary in Toronto.

We are,

Yours faithfully,

LL. WYNN JONES, Recorder,
CYRIL BURT, President (1923)
W. McDOUGALL, President (1924).

BOOK REVEIWS

ALBERT S. OSBORNE, *The Problem of Proof*, pp. xxii + 526. Matthew Bender & Co., New York and Albany, 1922.

Students of legal psychology should have access to this volume by the distinguished author of "Questioned Documents." Mr. Osborn has been a close student of human nature, especially as he has observed it in trial lawyers, judges, jurors, witness, litigants, and criminals. He understands perception and memory, reasoning and emotion, suggestion and persuasion, credulity and deceit; and he has here put down the rich accumulations of his wisdom for the guidance of trial attorneys and the furtherance of justice.

Of the twenty-eight chapters, the psychologist will especially enjoy these: Sifting the Evidence, The Atmosphere of a trial, Cross-Examination, Memory and the Proof of Facts, Advocacy, Persuasion and Practical Psychology in Courts of Law, Form Blindness, The Designing and Lighting of Court Rooms, and The Judge in Disputed Document Trials. The vocational counsellor and the prospective attorney should read, beginning on page 234, Mr. Osborn's description of the requirements of the great lawyer.

Carnegie Institute of Technology. W. V. BINGHAM

HELEN T. WOOLLEY and ELIZABETH FARRIS, *Diagnosis and Treatment of Young School Failures*, Bulletin 1923, No. 1, p. 115, 10 cents. Department of Interior, Bureau of Education, Washington, D. C.

Both the teacher and the psychologist will appreciate this picture of what may be accomplished with young school failures. Various mental and educational tests were used in selecting the sixteen children for the "demonstration class." The children in the class varied in chronological age from six years and seven months to ten years and four months, and the intelligence quotients of the first test varied from seventy-five to ninety-five.

The children were grouped for study into four dominant causes of their difficulties, as neglected children; high-grade defectives; special defects in some one subject; and psychopathic children. Each type is reported in the bulletin separately and each case is very completely reported, giving many helpful points in understanding such children.

The entire article is a plea for proper treatment and classification of exceptional children. One of the best arguments for Special Education that is shown in the bulletin is found in the discussion of the changed attitude of the children towards school and life. This is certainly worthy of consideration. Another is that the intelligence quotients of these children increased with the superior teaching.

Ohio University. M. LA VINIA WARNER

BERNARD C. EWER, *Applied Pschology* pp. xii+480. Macmillan Co., N. Y. 1923.

Psychology must be saved from her friends. Like many another well-favored jade of popular acclaim, she is in danger of having her name become a by-word and a term of reproach to be thrown in the teeth of her too much preoccupied guardians, the *zunft* of

Psychologists properly speaking. While they are engaged in cautious, painstaking effort to extend their own knowledge concerning her mysteries, interlopers and charlatans of all sorts are greedily capitalizing the credulity of an uniformed public, to the ultimate harm of both the public and the lady in question.

To show that she's not at all that kind of a gal is the *raison d'être* of the volume here under review. To give the author's own statement, "There is a lack of literature which attempts to present in a readable form the principles, methods, and results of scientific psychology as applied to problems of everyday life."

For those at all informed the author's thesis needs no more than its statement to be accepted. Book shops, magazine stands, popular lectures by "mentalists," all are more or less shamelessly vending under the name of Psychology various nostrums concocted at best under the animus of ignorant and misguided enthusiasm, at worst with cynical disregard for intellectual decency and with unscrupulous motives of gain. The method and techiques of science are caviar to the general, who as a rule prefer food of a different sort. Any effort such as the volume by Professor Ewer to cultivate a more discerning taste is wholly commendable. However much psychologists may differ as to details of method and interpretation—and that honest differences of opinion do exist no one will deny—there can be no disagreement on the general point of view espoused in this book, the point of view of any scientific discipline.

The book is divided into four parts. Part I gives a treatment of Aims, Principles, and Methods. After a brief survey of procedures and techniques as related to the aims of psychology a clarification of the concepts "subconscious," "unconscious," and "suggestion" are attempted. Part II deals with Education and Everyday Life—a sketch of what is ordinarily thought of as coming within the purview of educational psychology. Under Mind and Health, the general heading for Part III, the author treats those things which are probably most frequently associated with the term "psychology" for the layman—psychonalysis, psychotherapy in general, faith healing, etc. Part IV concerns itself with a survey of the methods and problems of psychology as applied to industry and commerce.

Probably any such popular introduction as this must be unsatisfactory from the standpoint of the special student. This would doubtless be true of sciences much better organized than psychology. A wise friend of the reviewer once defined all teaching as being a process of judicious lying. All of us who are engaged in the art necessarily have our individual notions as to what should be emphasized and what omitted. Since the book makes no pretense of being a compendium of all knowledge psychological, however, no strictures can properly be applied where the matter is still one of opinion among those qualified to judge, but only when questions of fact arise.

One rather serious lapse in this connection is the statement that the coefficient of correlation "in common use known as 'Pearson's coefficient of correlation is $\frac{6\text{sum}d^2}{n(n^2-1)}$." This formula is *not* in common use for serious correlation studies nor is it Pearson's formula.

A few other points on which psychologists might honestly differ may be briefly mentioned. (1) The history of the two terms "unconscious" and "subconscious" has not been such as to throw all the obloquy of representing a mythical potency rather than a scientific concept upon the former, as the author is inclined to do. The argument here reminds one of Mephistopheles' advise to the student in *Faust* to battle only with words.

"Denn eben wo Begriffe fehlen
Da stellt ein Wort zur rechten Zeit sich ein."

(2) "Subconsciousness works *mechanically*. It acts with a regularity of cause and effect which is distinct from rational self-direction and inhibition." Query, aside from the question of fact: is a *science* of psychology possible on any other basis than mechanistic postulates? (3) To subsume under the general heading of "special intelligence tests" such a diversity of tests as associates tests and special achievement tests of all sorts (educational tests) tends only further to befog an already very cloudy issue—especially for the lay person. (4) The attempted distinction between individual and "social" suggestion might be questioned. (5) "What is needed [in regard to religious psychotherapy] is.......a more studied practice of superposing religious faith upon medical care." (6) There is a *non sequitur* in the argument for the hereditary mental similarity of siblings (r of .40 to .75, which means 10 to 30 per cent better than a sheer guess) particularly when the author advances this argument: "Especially in twins is the mental similarity conspicuous, and since it is not substantially greater in older than in younger twins it cannot be ascribed to education." The relative weight of heredity *vs.* environment is not demonstrated in such figures, or by such argument. (7) In view of the fact that "Freudian psychology" is deprecated so generously—and some of the psychoanalytic literature certainly warrants this—it would seem to have been only fair to have differentiated more carefuly between the pronouncements of Freud and those of some of his over-enthusiastic followers. "The recommendation of its [Freudian doctrine] extreme advocates that one should refrain altogether from interfering with the emotions" is certainly counter to what Freud explicitly states in his General Introduction to Psychoanalysis—which the author lists in his bibliography.

With all these strictures the present writer feels that the book has a legitimate place. The intelligent layman and the beginning student should find it of value. Part IV on Industry and Commerce is particularly well done, and the chapter on Social and Moral Factors faces some of the larger problems involved in human adjustments.

Purdue University HERMANN REMMERS

DONALD A. LAIRD, Ph. D. *Applied Psychology for Nurses* pp. xi+236. J. B. Lippincott Co., Philadelphia, 1923.

In this little volume the author has "selected from the vast literature of psychology those facts of most immediate aid to nurses in understanding the patient, themselves, and their fellowmen," as announced on the paper cover. I should say that he has done more than that; he has gone further into the preserves of biology than many psychologists would venture—a procedure quite commendable, I believe.

Written in popular, newspaper English, the book should commend itself to those with the amount of preliminary training that the ordinary nurse has, as well as beginning students of psychology generally. The advanced student may be a bit irritated or even slightly scandalized at the irreverance manifested toward the household gods of psychology in some of the chapter and section headings—e.g., "Enter the Villain"; "A Practical Joke on a Sea-Anemone"; "The Moon and the Devil Blamed."

As already intimated the book is biological and behavioristic in viewpoint. It is divided into four parts. Part I attempts to orient the student with respect to the origin, scope, and applications of psychology, with particular reference to medicine and nursing; this is done in twenty-five pages. In Part II, the most technical part of the book, the author treats the biological aspects and foundations of behavior. Part III, entitled "Applications," is the *piece de resistance*: learning endicronology, temperaments, instincts, psychoanalysis, suggestion,—these, although not all called by their first names, indicate the rubrics under which the discussion may be classified. Part IV is given over to somewhat more general, sociological applications of psychology and psychiatry, with particular regard to the subject of mental hygiene.

The author states in the preface that he has "endeavored to avoid all controversial matter not borne out by fact," and has succeeded in the main. The chapter on The Temperaments in Nursing, however, offers opportunity for polemic to one controversially inclined. The ancient classification of the temperaments could well be improved from the modern point of view of abnormal psychology. To trace the inheritance of these various traits as unit characters (p. 126) seems to me more than questionable, the study of the Eugenics Record Office, which the author quotes, notwithstanding. Any mention of the glands of internal secretion upon behavior also opens up another debatable question. It is only fair to say, however, that the author treads rather lightly here.

These strictures are, after all, only minor and largely theoretical matters. There is sufficient warrant and need for a book of this sort to make it a valuable adjunct in the training of nurses. The portions of the book dealing with mental hygiene and the applications in general are well worthy the attention of the lay reader and beginning students generally. Unconventionally written as the book is, I am not prepared to state that this is a fault rather than a virtue.

Purdue University HERMANN REMMERS

MEAD, ARTHUR RAYMOND, *Learning and Teaching*. Lippencott Company, 1923. pp. XI and 277.

Learning and Teaching, as the title suggests, is a study in applied psychology. The author has endeavored to cause the student of education to apply the general laws of psychology to a great variety of class-room problems. He has consciously attempted to do something that will aid in removing the source of much criticism against present-day teaching of psychology and education in our modern teacher-training institutions. There is the persistent criticism, often very bitter, that instruction in psychology and education fails to carry over from the teacher-training institution to the actual teaching of boys and girls. It is this criticism that has prompted Professor Mead to offer the teachers of this country a new book on the subject.

Professor Mead has devoted one half of his 277 pages to exercises and references. These exercises and references cover a wide range of topics. Among the topics considered are the following: The nature of human learning, the general and individual factors that influence learning, the transfer of learning, the measurement of mental functions, psychological conditions concerned with the elementary and secondary curricula, and the psychological condition involved in educational methods. These topics receive a minimum of discussion. In fact the discussion of the text frequently is too limited to give the student adequate knowledge of the principle under consideration. Such information, however, may be readily gathered by the student from the excellent lists of references appended to each topic. This approach to the subject is familiar to those acquainted with the *Principles of Teaching* by Professor Thorndike.. The merits and limitations of such an approach are apparent.

Professor Mead has done students of education a real service by collecting this extended list of varied exercises and by carefully selecting references bearing on each topic. The exercises show discrimination in their selection and arrangement, and with the proper use of the suggested references, should give satisfactory results.

LOOSMORE, W. CHARLES. *The Gain of Personality.* E. P. Dutton & Co. pp. 233. Price $2.50.

The Gain of Personality is a popular psychological statement of the practical values of personality. In the course of his discussion, the author has treated the nature, elements, and general development of personality. Each of the three major divisions is analyzed into its component parts. Under the general head of element of personality we find such sub-divisions as individuality, sincerity, courage, enthusiasm, sensibility, self-expression, humor, and repose.

The book offers clear evidence that the author is acquainted with books and men. He states his position clearly and with effectiveness. Throughout he directs the reader's attention to those elements in the life of the individual necessary for success. He points out wherein many people fail, and indicates the elements necessary to success.

In brief the mission of the book is not to discuss personality "theoretically or philosophically" but "practically and helpfully." A reading of the book should aid the individual in taking stock of his personal merits and limitations.

WELLS, GEORGE ROSS. *Youth and the Open Door.* E. P. Dutton, 1922. pp. 175. Price $2.00.

The book contains a series of lectures originally delivered to freshmen in college. The purpose of the author is to present to young students certain laws of mental life whereby they may realize the largest benefits possible from their four years in college. With this point in mind, the author has discussed instinct, habit, remembering, attention, interest, thinking, emotions, causes of failure, choice of vocation, and character and temperament. The discussion of these topics is non-technical in character, and is designed to make effective certain laws of psychology. The author's psychology is modern and its application to the task in hand is clear and judicious.

The reader turns away from the book with the feeling that the effort of the author has been worthwhile. He has pointed out certain aids in the use of the mental powers that can not fail to benefit the student. It is certainly the type of book that should find its way into the hands of our youth who desire to make economic use of student days.

On the whole the book is well written and places its message before the reader in an effective manner.

In brief it is an effort to apply lessons of psychology to the very processes which make possible psychology.

HOKE, ROY EDWARD. *The Improvement of Speed and Accuracy in Typewriting.* The Johns Hopkins University Studies in Education, 1922. pp. 42. Price 75c.

Experimental studies in learning typewriting are not numerous. Book's well-known study concerned itself chiefly with the psychology of skill. He did not stress the improvement of speed, accuracy, or methods of teaching typewriting.

In his study, Dr. Hoke has selected five definite questions for investigation: First, an investigation of the frequency of the occurrence of letters and marks used in the English language; Second, an examination of the number and distribution of errors made in actual typewriting, and in so far as possible determine the causes of the errors; Third, a study of the relative abilities of the eight fingers and of the two hands; Fourth, the determination of the loads or burdens of work placed by the present system upon the fingers and hands; Fifth, an attempt to point out the consideration which must be admitted in an effort to arrange a keyboard on scientific principles.

The study closes with a suggested keyboard together with certain criticism of the proposed changes in the keyboard. There is appended a bibliography of twenty-six titles. As would be expected most of the titles discuss the more general problem of the acquisition of skill.

To the reviewer the study deserves careful consideration by those interested in teaching typewriting and who are anxious to produce a product with the highest skill and accuracy. It at least suggests a method of approach of investigating ways and means of improving skill in one of the most widely used skills of man.

HOKE, ELMER RHODES. *The Measurement of Achievement in Shorthand.* The Johns Hopkins University Studies in Education, 1922. pp. 118. Price $1.50.

In this monograph Professor Hoke gives an account of four scales that he has constructed for use in connection with teaching Gregg Shorthand. The author recognizes the fact that the use of shorthand is a complex process, involving mental and motor factors. In his analysis of the process, the author selects four important aspects for consideration. These important aspects are reading ability, quality of writing, speed of writing, and knowledge of the system. A separate scale has been constructed for each of the four items. Samples of each scale are printed in the appendix of the monograph. Tentative norms for each have been determined. It is the authors belief, and no doubt with a great deal of justification, that such scales in the hands of teachers of shorthand will result in better methods of teaching and consequently better qualified users of shorthand.

WILLIS L. GARD

Ohio University

NEW BOOKS AND PAMPHLETS RECEIVED*

Books and pamphlets for review should be sent to James P. Porter, Department of Psychology, Ohio University, Athens, Ohio.

BAUDOUIN, L. CHARLES. (translated by Fred Rothwell.) *The Birth of Psyche.* E. P. Dutton & Company, New York City. 211 pp.

BYRNE, LEE. *Latin Tests in Iowa High Schools.* University of Iowa, Iowa City, Iowa. 40 pp.

DICKSON, VIRGIL E. *Mental Tests and the Classroom Teacher.* World Book Company, Yonkers-on-Hudson, N. Y. 231 pp.

EWER, BERNARD C. *Applied Psychology,* Macmillan Company, New York. 480 pp.

FORBES, JOSEPH. *Lehrbuch der Experimentellen Psychologie.* Herder & Co. G. m. b. H. Verlagsbuchhandlung, Freiburg im Breisgau. 631 pp.

GATES, ARTHUR I. *Psychology for Students of Education.* The Macmillan Company, New York. 489 pp.

HALBERT, L. A. *What is Professional Social Work?* The Survey, 112 East 19th St., New York City. 149 pp.

HALL, G. STANLEY. *Life and Confessions of a Psychologist.* D. Appleton Company, New York. Price $5.00. 623 pp.

HENDERSON, C. A. *Personal and Business Efficiency.* George H. Doran Company, New York. Price $2.00. 308 pp.

MONSARRAT, K. W. *Health and the Human Spirit.* E. P. Dutton & Company, New York. Price $2.00. 132 pp.

PUBLICATIONS ISSUED BY THE DEPARTMENT OF INTERIOR, BUREAU OF EDUCATION, WASHINGTON, D. C.

Americanization in the United States. John J. Mahoney. Bulletin No. 31, 1923. 42 pp.

Analytic Survey of State Courses of Study for Rural Elementary Schools. Charles M. Reinoehl. Bulletin No. 42, 1922. 116 pp.

An Americanization Program. E. J. Irwin. Bulletin No. 30, 1923. 60 pp.

Art Educations: The Present Situation. Royal Bailey Farnum. Bulletin No. 13, 1923. 20 pp.

Athletic Badge Tests for Boys and Girls. Physical Education Series No. 2, 1923. Prepared by the Playground and Recreation Association of America. 17 pp.

Bible in the Public Schools. William R. Hood. Bulletin No. 15, 1923. 13 pp.

Child Health School. Conducted at the University of Chicago, 1920. Lydia J. Roberts. School Health Studies. No. 2, 1923. 60 pp.

College Entrance Credits in Commercial Subjects. Glen Levin Swiggett. Commercial Education Leaflet No. 4, April, 1923. 22 pp.

---

*Mention here does not preclude further comment.

Consolidation of Schools in Randolph County, Indiana. O. H. Greist. Rural School Leaflet No. 12, 1923. 11 pp.

Consolidation of Schools of Weld County, Colorado. C. G. Sargent. Rural School Leaflet No. 13, 1923. 11 pp.

Educational Extension. Charles G. Maphis. Bulletin No. 24, 1923. 32 pp.

Educational Hygiene. Willard S. Small. Bulletin No. 33, 1923. 36 pp.

Educational Progress and the Parents. Orville G. Brim Rural School Leaflet No. 15, May, 1923. 8 pp.

Educational Surveys. Edward Franklin Buchner. Bulletin No. 17, 1923. 44 pp.

Educational Work of The Knights of Columbus. Mark J. Sweeny. Bulletin No. 22, 1923. 12 pp.

Educational Work of the Young Women's Christian Association. By Education and Research Division, National Board of Y. W. C. A. Bulletin No. 26, 1923. 24 pp.

Enrollment in Foreign Languages, Science and Commercial Subjects in Public High Schools. Florence DuBois. Statistical Circular No. 2, May, 1923. 9 pp.

Federal Aid to Public Schools. Fletcher Harper Swift. Bulletin No. 47, 1922. 47 pp.

Hampton Normal and Agricultural Institute. Walton C. John. Bulletin No. 27, 1923. 116 pp.

Higher Education, 1920-1922. George F. Zook. Bulletin No. 34, 1923. 33 pp.

History of the Manual Training School of Washington University. (St. Louis Training School.) Charles Penney Coates. Bulletin No. 3, 1923. 86 pp.

Home Economics Education. Henrietta W. Calvin. Bulletin No. 6, 1923. 19 pp.

Iowa Plan of Training Superintendents and Teachers for Consolidated Schools. Macy Campbell. Rural School Leaflet No. 17, June, 1923.

Junior High Schools of Berkeley, California. James T. Preston and others. Bulletin No. 4, 1923. 48 pp.

Main Streets of the Nation. A Series of Projects on Highway Transport for Elementary Schools. Florence C. Fox. Bulletin No. 38, 1923. 42 pp.

Medical Education, 1920-1922. N. P. Colwell, M. C. Bulletin No. 18, 1923. 17 pp.

Objectives in Elementary Rural School Agriculture. E. E. Windes. Rural School Leaflet No. 11, March, 1923. 18 pp.

Organization and Conduct of Business. Glen Levin Swiggett. Commercial Education Leaflet No. 5, June, 1923. 18 pp.

Principles of Kindergarten-Primary Education in the Consolidated Rural School. Katherine M. Cook. Rural School Leaflet No. 18, June, 1923. 9 pp.

Progressive Tendencies in European Education. C. W. Washburne. Bulletin No. 37, 1923. 31 pp.

Public Education in Oklahoma. Bulletin No. 14, 1923. 92 pp.

Public School System of Arkansas. Bulletin No. 11, 1923. 110 pp.

Recent Advances in Instruction in Music. Will Earhart and Charles N. Boyd. Bulletin No. 30, 1923. 21 pp.

Recent Developments in Educational Journalism. W. Carson Ryan, Jr. Bulletin No. 25, 1923. 14 pp.

Record of Current Educational Publications. Bulletin No. 32, 1923. 36 pp.

References on Education of Women in the United States. Library Leaflet No. 19, April, 1923. 7 pp.

References on Preschool and Kindergarten-Primary Education. Nina C. Vandewalker and Harriet E. Howard. Kindergarten Circular No. 14, July, 1923. Price per copy 5 cents. 11 pp.

Rural-Teacher Situation in the United States. Mabel Carney. Rural School Leaflet No. 14, May, 1923. Price per copy 5 ~~cents.~~ 12 pp.

School Building Program for Washington, North Carolina. Alice Barrows. Bulletin No. 2, 1923. 20 pp.

School Support and School Indebtedness in Cities. Florence DuBois. Statistical Circular No. 3, June, 1923. 5 pp.

Social Studies in Civic Education. Edgar Dawson. Bulletin No. 23, 1923. 16 pp.

Some Industrial Art Schools of Europe. Marius Vachon. Translation by Florence N. Levy. Bulletin No. 48, 1922. 59 pp.

State Policies in Public School Finance. Fletcher Harper Swift. Bulletin No. 6, 1922. 54 pp.

Statistical Survey of Education, 1919-1920. Florence DuBois. Bulletin No. 16, 1923. 41 pp.

Statistics of City School Systems, 1919-1920. H. R. Bonner. Bulletin No. 17, 1922. 173 pp.

Statistics Relating to Business Education in Colleges and Universities, 1921-1922. Glen Levin Swiggett. Commercial Education Leaflet No. 6, June, 1923. 8 pp.

Supervision of One-Teacher Schools. Maud C. Newbury. Bulletin No. 9, 1923. 55 pp.

Vocational Education. William T. Bawden. Bulletin No. 28, 1923. 26 pp.

Robinson, George Edwin. *Scientific Salesmanship.* Three Volumes. American College of Scientific Salesmanship, Chicago, Ill.

Skaggs, Ernest Burton. *Elementary Text Book of Mental Measurements.* George Wahr, Publisher, Ann Arbor, Mich. 169 pp.

Starch, Daniel. *Principles of Advertising.* A. W. Shaw Company, New York. 998 pp.

Wates, George Frederick. *The Magic of Common Sense.* E. P. Dutton & Company, New York. Price $1.60. 107 pp.

Wood, Ben D. *Measurement in Higher Education.* World Book Company, Yonkers-on-Hudson, N. Y. 337 pp.

INDEX OF SUBJECTS

(The subject headings of all contributions except Book Reviews and Notes are printed in SMALL CAPITALS.)

Achievement of Subnormal Children in Standardized Educational Tests (Wallin) 108
ADVANCED COURSE IN EXPERIMENTAL PSYCHOLOGY APPLIED TO ADVERTISING 360
ADVERTISEMENTS, FACIAL EXPRESSIONS IN 114
ADVERTISING, CONDITIONS OF BELIEF IN 1; EXPERIMENTAL PSYCHOLOGY APPLIED TO, ADVANCED COURSE IN 360·
MINOR STUDIES IN PSYCHOLOGY OF 10
ADVERTISING EFFICIENCY, RETURN COUPON AS A MEASURE OF 202
AGRICULTURAL ENGINEERS, VOCATIONAL TESTS FOR 178
APTITUDE, ENGINE LATHE 16
APTITUDE TESTS, PREDICTION FORMULAE FOR TEAMS OF 277
ARMY TRADE TESTS RE-INTERPRETATION OF THE STATISTICAL METHOD OF 339
ATTENTION, FRANZ DOT TAPPING TEST AS A MEASURE OF 353
BASIS OF TOOTH PASTE SALES IN REPRESENTATIVE COMMUNITIES 173
BASKERVILLE ROMAN, BODONI AND CHELTENHAM TYPE FACES, LEGIBILITY OF 237
Behaviorism and Psychology (Roback) 192
BELIEF IN ADVERTISING, CONDITIONS OF 1
Beliefs, Expressed by Applied Psychology Students 102; Note on Popular Pseudo-Psychological 101
BINET TESTS, THORNDIKE COLLEGE ENTRANCE TESTS AND FIRST SEMESTER GRADES, THE FRESHMAN: 77
BODONI, BASKERVILLE ROMAN AND CHELTENHAM TYPE FACES, LEGIBILITY OF 237
BOYS' GANGS, INFLUENCE OF MENTAL LEVEL IN FORMATION OF 224
Causative Factors in Mental Inferiority and the Prevention of Degeneracy (Wallin) 294
CHELTENHAM TYPE FACES, BODONI AND BASKERVILLE ROMAN, LEGIBILITY OF 237
CHILDREN, SPECIAL SCHOOLS, PERIOD OF EIGHT YEARS, REPEATED MENTAL EXAMINATION OF 200 DEFECTIVE 54
CHOICE OF HIGH SCHOOL PUPILS, RELATION BETWEEN THE INTELLIGENCE AND VOCATIONAL 143
CLINICAL USE, COMBINATION MENTAL TEST FOR 246
COLLEGE ENTRANCE TESTS, FIRST SEMESTER GRADES, BINET TESTS, THE FRESHMAN: THORNDIKE 77
COLLEGE, FRESHMAN TESTS IN THE SMALL 258; INTELLIGENCE TESTS VERSUS ENTRANCE EXAMINATIONS, A MEANS OF PREDICTING SUCCESS IN 330
COLLEGE STUDENTS' ESTIMATES OF INTELLIGENCE WITH OTIS TESTS AND OTHER SCALES, CORRELATION OF 157
COMBINATION MENTAL TEST FOR CLINICAL USE 246
COMPARATIVE SOCIAL TRAITS OF VARIOUS RACES 127
Comparison of Intelligence and Training of School Children in a Massachusetts Town (Shaw & Lincoln) 296
CONDITIONS OF BELIEF IN ADVERTISING 1
CONTRIBUTION TO THE TECHNIQUE OF PARTIAL CORRELATION 135

Correlation, Technique of Partial 135; of College *Students'* Estimates of Intelligence With Otis Tests 157
Defective Children. Repeated Mental Examinations of 200 54
Deficiency With Musical Talent, Case of Secondary Mental 349
Divining Experiments on Rod- 303
Dot Tapping Test, a Measure of Attention, Franz 353
Downey Will-Temperament Group Test, Reliability and Validity of 65
Downey Will-Temperament Test. Racial Differences as Measured by 30
Educational Conference, College of Education, Ohio State University 188
Efficiency, Return Coupon as Measure of Advertising 202
Elements of Scientific Psychology (Dunlap) 189
Engineers, Vocational Tests for Agricultural 178
Engine Lathe Aptitude, Experiment in Testing 16
Entrance Examinations Versus Intelligence Tests as Means of Predicting Success in College 330
Estimates of Intelligence With Otis Tests and Other Scales, Correlation of College Student's 157
Examinations of 200 Defective Children Attending Special Schools, Repeated Mental 54
Experimental Psychology Applied to Advertising, Course in 360
Experiments in Testing Lathe Engineer Aptitude 16; on Rod-Divining 303
Facial Expressions in Advertising 111
Fluctuation in Mental Production, Motivation the Main Variable 209
Formulae for Teams of Aptitude Tests, Prediction 277
Franz Dot Tapping Test, Measure of Attention 353
Freshman Tests in the Small College 258
Freshman: Thorndike College Entrance Tests, First Semester Grades, Binet Tests, the 77
Gangs, Mental Level in the Formation of Boys' 224
General Problems of Psychology (MacDougall) 191
Group Tests of Mental Ability, *Reliability of* M.A. and IQ, Based on 93; Reliability and Validity, Downey Will Temperament 65
Health Fellowships and Scholarships 291
High Schools and Sex Education 109
High School *Pupils*, the Intelligence and Vocational Choices of 143
Honesty Book 198
How Numerals are Read (Terry) 106
Human Instincts in Business (Fitt) 297
Incomplete Rankings *Scheme for Combining* 168
Influence of Mental Level in Formation of Boys' Gangs 224
IQ, Based on Group Test of General Mental Ability, Reliability of M.A. and 93
Intelligence With Otis and Other Scales, Correlation of College Students' Estimate of 157

INDEX OF SUBJECTS

INTELLIGENCE TESTS VERSUS ENTRANCE EXAMINATIONS AS MEANS OF PREDICTING SUCCESS IN COLLEGE 330
INTELLIGENCE AND VOCATIONAL CHOICES OF HIGH SCHOOL PUPILS, RELATION BETWEEN 143
Interrelation of Some Higher Learning Processes (Haught) 193
Investigation of the Sex Relationship, Marriage, Delinquency and Truancy of Children Assigned to Special Public School Classes (Wallin) 294
Job Analysis and the Curriculum (Strong and Uhrbrock) 295
Junior High School Mathematics (Pickell) 195
LABOR TURNOVER, MENTAL ALERTNESS TEST SCORES AND 285
LEGIBILITY OF BODONI, BASKERVILLE ROMAN AND CHELTENHAM TYPE FACES 237
MEASURE OF ADVERTISING EFFICIENCY, RETURN COUPON AS 202; OF ATTENTION, FRANZ DOT TAPPING TEST 353
MEASURING THE COMPREHENSION OF ADVERTISEMENTS 364
MENTAL ABILITY, RELIABILITY M.A. AND IQ BASED ON GROUP TESTS;– AGE AND IQ, BASED ON GROUP TESTS OF GENERAL MENTAL ABILITY, THE RELIABILITY OF 93; –ALERTNESS TEST SCORE, POSITIONS AND PERMANENCY IN A COMPANY, RELATION OF 154; –ALERTNESS TEST SCORES, LABOR TURNOVER AND 285;– DEFICIENCY, WITH MUSICAL TALENT, SECONDARY 349; –EXAMINATIONS, 200 DEFECTIVE CHILDREN, SPECIAL SCHOOLS, REPEATED 54; LEVEL, FORMATION OF... BOYS' GANGS, INFLUENCE OF 224; TEST FOR CLINICAL USE, COMBINATION 246
Methods of Vocational Guidance (Claparede) 103
MINOR STUDIES IN THE PSYCHOLOGY OF ADVERTISING 10
MOTIVATION THE MAIN VARIABLE, FLUCTUATIONS IN MENTAL PRODUCTION WHEN 209
MUSICAL TALENT, SECONDARY MENTAL DEFICIENCY WITH 349
PARTIAL CORRELATION, TECHNIQUE OF 135
Percy Bysshe Shelley (Moore) 196
Poetic Mind (Prescott) 104
POSITIONS AND PERMANENCY IN A COMPANY, RELATION OF MENTAL ALERTNESS TEST SCORE 154
PREDICTION FORMULAE FOR TEAMS OF APTITUDE TESTS 277
Professional Education of Teachers, Cleveland 193
Pseudo-Psychological Beliefs, Note on 101
PSYCHOLOGY OF ADVERTISING, MINOR STUDIES 10
Psychology of Reading and Spelling (Gates) 193
Qualities Related to Success in Teaching (Knight) 107
RACES, SOCIAL TRAITS OF VARIOUS 127
RACIAL DIFFERENCES MEASURED BY DOWNEY WILL-TEMPERAMENT TEST 30
RANKINGS, SCHEME FOR COMBINING INCOMPLETE 168
RE-INTERPRETATION OF THE STATISTICAL METHOD OF ARMY TRADE TESTS 339
RELATION BETWEEN THE INTELLIGENCE AND VOCATIONAL CHOICES OF HIGH SCHOOL PUPILS 143; OF MENTAL ALERTNESS TEST SCORES TO POSITIONS AND PERMANENCY IN A COMPANY 154
RELIABILITY, M. A. AND IQ ON GROUP TEST OF GENERAL MENTAL ABILITY 93; AND VALIDITY DOWNEY

INDEX OF AUTHORS

(The names of authors of original contributions are printed in SMALL CAPITALS.)

ARLITT, ADA HART 330
ANDERSON, META L. 54

BARNES, C. W. 10
BASCH, CORYNE 237
BILLS, MARIAN A. 164
BURTT, HAROLD E. 114, 178, 237

CLARK, J. CAMDEN 114
CRAYTOR, LAURA C. 127

DASHIELL, J. F. 30
DAVENPORT, C. B. 127
DEL MANZO, M. C. 65

FEINGOLD, GUSTAVE A. 143
FOSTER, WILLIAM S. 308
FRANKEN, R. B. 312

Gard, W. L 196
GATES, ARTHUR I. 93
Good, H. G. 194, 297
Growdon, C. H. 109, 191, 300

HAUGHT, B. F. 168
HUFFAKER, C. L. 135
HULL, CLARK L. 277

IUPPENLATZ, C. 10
IVES, F. U. 178

JEFFERSON, B. L. 104

Kantor, J. R. 192
KENT, GRACE H. 246
KITSON, H. D. 10, 103
KNIGHT, F. B. 209
KORNHAUSER, ARTHUR W. 339

LAIRD, DONALD A. 173

MCFADDEN, JOHN F. 30
MINOGUE, BLANCHE M. 349
Morton, R. L. 106
MURRAY, ELSIE 258

PATTEN, EVERETT F. 16
Patterson, Donald G. 102
POFFENBERGER, A. T.
1, 202, 303, 304

Porter, James P. 198

REMMERS, HERMAN H. 107, 209
ROOT, W. T. 77
RUCH, G. M. 65
SHERMAN, IRENE CASE 353
SNOW, A. J. 285, 296, 298, 360
WARNER, M. LA VINIA 108.
196, 224, 294, 295

Wilson, Hiram R. 197
YARBROUGH, J. U. 157